# Dynamical Behavior of a Stochastic Forward-Backward Algorithm Using Random Monotone Operators

Pascal Bianchi *  Walid Hachem†

July 5, 2016

**Abstract**

The purpose of this paper is to study the dynamical behavior of the sequence produced by a forward-backward algorithm, involving two random maximal monotone operators and a sequence of decreasing step sizes. Defining a mean monotone operator as an Aumann integral, and assuming that the sum of the two mean operators is maximal (sufficient maximality conditions are provided), it is shown that with probability one, the interpolated process obtained from the iterates is an asymptotic pseudo trajectory in the sense of Benaïm and Hirsch of the differential inclusion involving the sum of the mean operators. The convergence of the empirical means of the iterates towards a zero of the sum of the mean operators is shown, as well as the convergence of the sequence itself to such a zero under a demipositivity assumption. These results find applications in a wide range of optimization problems or variational inequalities in random environments.

**Keywords :** Dynamical systems, Random maximal monotone operators, Stochastic forward-backward algorithm, Stochastic proximal point algorithm.

**AMS subject classification :** 47H05, 47N10, 62L20, 34A60.

## 1 Introduction

In the fields of convex analysis and monotone operator theory, the forward-backward splitting algorithm [1, 2] is one of the most often studied techniques for iteratively finding a zero of a sum of two maximal monotone

*LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay. e-mail: pascal.bianchi@telecom-paristech.fr

†LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay. e-mail: walid.hachem@telecom-paristech.fr

operators. This problem finds applications in convex minimization problems. Indeed, when each of the two maximal monotone operators coincides with the subdifferential of a proper and lower semicontinuous convex function, the forward-backward algorithm converges to a minimizer of the sum of the two functions, provided some conditions are met. Other applications include saddle point problems and variational inequalities. Each iteration of the algorithm involves a forward step, where one of the operators is used explicitly, followed a backward step that consists in applying the *resolvent* of the second operator to the output of the forward step.

The purpose of this paper is to study a version of the forward-backward algorithm, where at each iteration, each of the two operators is replaced with an operator that has been *randomly chosen* amongst a collection of maximal monotone operators. The sequence of random monotone operators is assumed to be independent and identically distributed (in a sense that will be made clear below), and the step size of the algorithm is supposed to approach zero as the number of iterations goes to infinity, in order to alleviate the noise effect due to the randomness.

The aim is to study the dynamical behavior of the stochastic sequence generated by the above algorithm. Our main result states that the piecewise linear interpolation of the output sequence is an *asymptotic pseudo-trajectory* (APT) [3, 4] of a certain *semiflow*, which we shall characterize below. Loosely speaking, it means that the iterates of our stochastic forward-backward algorithm asymptotically "shadow" the trajectory of a continuous time dynamical system, hence inheriting its convergence properties. In our case, the latter dynamical system is taken as a differential inclusion involving the sum of the *Aumann expectations* of the randomly chosen maximal monotone operators [5, 6], as also introduced in the recent paper [7].

The convergence of the algorithm towards an element of the set of zeros of the sum of the Aumann expectations is of obvious interest. In this regard, the above APT property yields two important corollaries. Using a result of [8], we show that the sequence of empirical means of the iterates converges almost surely (a.s.) to a (random) element of the set of zeros. Moreover, when the sum of the Aumann expectations is assumed *demipositive* [9], we prove that the sequence of iterates converges a.s. to a zero. Verifiable conditions for demipositivity can be easily devised.

This paper is organized as follows. Section 2 provides the theoretical background. Section 3 introduces the main algorithm and states the main results. Section 4 reviews some applications to convex minimization problems. Related works are discussed in Section 5. Proofs are provided in Section 6. Perspectives and conclusions are addressed in Sections 7 and 8 respectively.

## 2 Preliminaries

### 2.1 Monotone Operators

A set-valued operator $\mathsf{A} : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$, where $N$ is some positive integer, is said to be monotone if $\forall (x, y) \in \mathrm{gr}(\mathsf{A})$, $\forall (x', y') \in \mathrm{gr}(\mathsf{A})$, $\langle y - y', x - x' \rangle \geq 0$, where $\mathrm{gr}(\mathsf{A})$ stands for the graph of $\mathsf{A}$. A non-empty monotone operator is said to be maximal if its graph is a maximal element in the inclusion ordering. A typical maximal monotone operator is the subdifferential of a function belonging to $\Gamma_0$, the family of proper and lower semicontinuous convex functions on $\mathbb{R}^N$. We use $\mathcal{M}$ to represent the set of maximal monotone operators on $\mathbb{R}^N$, and let $\mathrm{dom}(\mathsf{A}) := \{x \in \mathbb{R}^N : \mathsf{A}(x) \neq \emptyset\}$ be the domain of the operator $\mathsf{A}$.

Given that $\mathsf{A}, \mathsf{B} \in \mathcal{M}$, where $\mathsf{B}$ is assumed to be single-valued and where $\mathrm{dom}(\mathsf{B}) = \mathbb{R}^N$, the forward-backward algorithm reads

$$\mathsf{x}_{n+1} = (I + \gamma \mathsf{A})^{-1}(\mathsf{x}_n - \gamma \mathsf{B}(\mathsf{x}_n)), \tag{1}$$

where $I$ is the identity operator, $\gamma$ is a real positive step, and $(\cdot)^{-1}$ is the inverse operator defined by the fact that $(x, y) \in \mathrm{gr}(A^{-1}) \Leftrightarrow (y, x) \in \mathrm{gr}(A)$ for an operator $A$. The operator $(I + \gamma \mathsf{A})^{-1}$, called the resolvent, is single valued with the domain $\mathbb{R}^N$ since $\mathsf{A} \in \mathcal{M}$ [10, 11]. In the special case where $\mathsf{A}$ is equal to the subdifferential $\partial f$ of a function $f \in \Gamma_0$, the resolvent is also refered to as the proximity operator, and we note $\mathrm{prox}_f(x) = (I + \partial f)^{-1}(x)$.

We denote the set of zeros of $\mathsf{A}$ as $Z(\mathsf{A}) := \{x \in \mathbb{R}^N : 0 \in \mathsf{A}(x)\}$. Assuming that $\mathsf{B}$ is so-called cocoercive, and that $\gamma$ satisfies a certain condition, the forward-backward algorithm is known to converge to an element of $Z(\mathsf{A} + \mathsf{B})$, provided the latter set is not empty [11].

### 2.2 Set-Valued Functions and Set-Valued Integrals

Let $(\Xi, \mathscr{T}, \mu)$ be a probability space, where $\mathscr{T}$ is $\mu$-complete. Consider the space $\mathbb{R}^N$ equipped with its Borel field $\mathscr{B}(\mathbb{R}^N)$, and let $F : \Xi \rightrightarrows \mathbb{R}^N$ be a set-valued function such that $F(\xi)$ is a closed set for any $\xi \in \Xi$. The set-valued function $F$ is said to be *measurable* if $\{\xi : F(\xi) \cap H \neq \emptyset\} \in \mathscr{T}$ for any set $H \in \mathscr{B}(\mathbb{R}^N)$. This is known to be equivalent to asserting that the domain $\mathrm{dom}(F) := \{\xi \in \Xi : F(\xi) \neq \emptyset\}$ of $F$ belongs to $\mathscr{T}$, and that there exists a sequence of measurable functions $\varphi_n : \mathrm{dom}(F) \to \mathbb{R}^N$ such that $F(\xi) = \mathrm{cl}(\{\varphi_n(\xi)\})$ for all $\xi \in \mathrm{dom}(F)$ [12, Chap. 3] [13]. Assume now that $F$ is measurable and that $\mu(\mathrm{dom}(F)) = 1$. For $1 \leq p < \infty$, let $\mathcal{L}^p(\Xi, \mathscr{T}, \mu; \mathbb{R}^N)$ be the Banach space of measurable functions $\varphi : \Xi \to \mathbb{R}^N$ with $\int \|\varphi\|^p d\mu < \infty$, and let

$$\mathcal{S}_F^p := \{\varphi \in \mathcal{L}^p(\Xi, \mathscr{T}, \mu; \mathbb{R}^N) : \varphi(\xi) \in F(\xi) \ \mu - \text{a.e.}\}. \tag{2}$$

If $\mathcal{S}_F^1 \neq \emptyset$, then the function $F$ is said to be integrable. The *Aumann integral* [5, 6] of $F$ is the set

$$\int F d\mu := \left\{ \int_\Xi \varphi d\mu \ : \ \varphi \in \mathcal{S}_F^1 \right\} .$$

### 2.3 Random Maximal Monotone Operators

Consider the function $A : \Xi \to \mathcal{M}$. Note that the graph $\mathrm{gr}(A(\xi, \cdot))$ of any element $A(\xi, \cdot)$ is a closed subset of $\mathbb{R}^N \times \mathbb{R}^N$ by the maximality of $A(\xi, \cdot)$ [10, Prop. 2.5]. Assume that the function $\xi \mapsto \mathrm{gr}(A(\xi, \cdot))$ is measurable as a closed set-valued $\Xi \rightrightarrows \mathbb{R}^N \times \mathbb{R}^N$ function. It is shown in [14, Ch. 2] that this is equivalent to saying that the function $\xi \mapsto (I + \gamma A(\xi, \cdot))^{-1} x$ is measurable from $\Xi$ to $\mathbb{R}^N$ for any $\gamma > 0$ and any $x \in \mathbb{R}^N$. If the domain of $A(\xi, \cdot)$ is represented by $D(\xi)$, the measurability of $\xi \mapsto \mathrm{gr}(A(\xi, \cdot))$ implies that the set-valued function $\xi \mapsto \mathrm{cl}(D(\xi))$ is measurable. Moreover, recalling that $A(\xi, x)$ is the image of a given $x \in \mathbb{R}^N$ under the operator $A(\xi, \cdot)$, the set-valued function $\xi \mapsto A(\xi, x)$ is measurable [14, Ch. 2]. Given $x \in D(\xi)$, the element of least norm in $A(\xi, x)$ is denoted as $A_0(\xi, x)$. In other words, $A_0(\xi, x) = \mathrm{proj}_{A(\xi,x)}(0)$. It is known that the function $\xi \mapsto A_0(\xi, x)$ is measurable [14, Ch. 2].

For any $\gamma > 0$, the resolvent of $A(\xi, \cdot)$ is represented by

$$J_\gamma(\xi, x) := (I + \gamma A(\xi, \cdot))^{-1}(x).$$

As we know, $J_\gamma(\xi, \cdot)$ is a non-expansive function on $\mathbb{R}^N$. Since $J_\gamma(\xi, x)$ is measurable in $\xi$ and continuous in $x$, Carathéodory's theorem shows that the function $J_\gamma : \Xi \times \mathbb{R}^N \to \mathbb{R}^N$ is $\mathscr{T} \otimes \mathscr{B}(\mathbb{R}^N)$ measurable. We also introduce the Yosida approximation $A_\gamma(\xi, \cdot)$ of $A(\xi, \cdot)$, which is defined for any $\gamma > 0$ as the $\mathscr{T} \otimes \mathscr{B}(\mathbb{R}^N)$ measurable function

$$A_\gamma(\xi, x) := \frac{x - J_\gamma(\xi, x)}{\gamma} .$$

The function $A_\gamma(\xi, \cdot)$ is a $\gamma^{-1}$-Lipschitz continuous function that satisfies $\|A_\gamma(\xi, x)\| \uparrow \|A_0(\xi, x)\|$ and $A_\gamma(\xi, x) \to A_0(\xi, x)$ for any $x \in D(\xi)$ when $\gamma \downarrow 0$. Moreover, the inclusion $A_\gamma(\xi, x) \in A(\xi, J_\gamma(\xi, x))$ holds true for all $x \in \mathbb{R}^N$ [10, 11].

The *essential intersection* $\mathcal{D}$ of the domains $D(\xi)$ is [15]

$$\mathcal{D} := \bigcup_{E \in \mathscr{T} : \mu(E) = 0} \ \bigcap_{\xi \in \Xi \setminus E} D(\xi) ,$$

in other words, $x \in \mathcal{D} \ \Leftrightarrow \ \mu(\{\xi : x \in D(\xi)\}) = 1$. Let us assume that $\mathcal{D} \neq \emptyset$ and that this function is integrable for each $x \in \mathcal{D}$. On $\mathcal{D}$, we define $\mathcal{A}$ as the Aumann integral

$$\mathcal{A}(x) := \int_\Xi A(\xi, x) \mu(d\xi) .$$

One can immediately see that the operator $\mathcal{A} : \mathcal{D} \rightrightarrows \mathbb{R}^N$ so defined is a monotone operator.

### 2.4 Evolution Equations and Almost Sure APT

Given that $\mathsf{A} \in \mathcal{M}$, consider the differential inclusion

$$\dot{z}(t) \in -\mathsf{A}(z(t)) \quad \text{a.e. on } \mathbb{R}_+, \quad z(0) = z_0, \tag{3}$$

for a given $z_0$ in $\mathrm{dom}(\mathsf{A})$. It is known from [10, 16] that for any $z_0 \in \mathrm{dom}(\mathsf{A})$, there exists a unique absolutely continuous function $z : \mathbb{R}_+ \to \mathbb{R}^N$ satisfying (3) - referred to as the *solution* to (3). Consider the map

$$\Psi : \mathrm{dom}(\mathsf{A}) \times \mathbb{R}_+ \to \mathrm{dom}(\mathsf{A}), \quad (z_0, t) \mapsto z(t),$$

where $z(t)$ is the solution to (3) with the initial value $z_0$. Then, for any $t \geq 0$, $\Psi(\cdot, t)$ is a non-expansive map from $\mathrm{dom}(\mathsf{A})$ to $\mathrm{dom}(\mathsf{A})$ who can be extended by continuity to a non-expansive map from $\mathrm{cl}(\mathrm{dom}(\mathsf{A}))$ to $\mathrm{cl}(\mathrm{dom}(\mathsf{A}))$ that we still denote as $\Psi(\cdot, t)$ [10, 16]. The function $\Psi$ so defined is a *semiflow* on the set $\mathrm{cl}(\mathrm{dom}(\mathsf{A})) \times \mathbb{R}_+$, being a continuous function from $\mathrm{cl}(\mathrm{dom}(\mathsf{A})) \times \mathbb{R}_+$ to $\mathrm{cl}(\mathrm{dom}(\mathsf{A}))$, satisfying $\Psi(\cdot, 0) = I$ and $\Psi(z_0, t+s) = \Psi(\Psi(z_0, s), t)$ for every $z_0 \in \mathrm{cl}(\mathrm{dom}(\mathsf{A}))$, $t, s \geq 0$. The set $\gamma(x) := \{\Psi(x, t) : t \geq 0\}$ is the *orbit* of $x$. Although orbits of $\Psi$ are not necessarily convergent in general, any solution to (3) converges to a zero of $\mathsf{A}$ (which is assumed to exist) whenever $\mathsf{A}$ is *demipositive* [9]. By demipositive, we mean that there exists $w \in Z(\mathsf{A})$ such that for every sequence $((u_n, v_n) \in \mathsf{A})$ such that $(u_n)$ converges to $u$ and $\{v_n\}$ is bounded,

$$\langle u_n - w, v_n \rangle \xrightarrow[n\to\infty]{} 0 \quad \Rightarrow \quad u \in Z(\mathsf{A}).$$

We now need to introduce some important notions associated with the semiflow $\Psi$. A comprehensive treatment of the subject can be found in [3, 17]. A set $S \subset \mathrm{cl}(\mathrm{dom}(\mathsf{A}))$ is said to be *invariant* for the semiflow $\Psi$ if $\Psi(S, t) = S$ for all $t \geq 0$. Given that $\varepsilon > 0$ and $T > 0$, a $(\varepsilon, T)$*-pseudo orbit* from a point $a$ to a point $b$ in $\mathbb{R}^N$ is a $n$-uple of partial orbits $(\{\Psi(y_i, s) : s \in [0, t_i]\})_{i=0,\ldots,n-1}$ such that $t_i \geq T$ for $i = 0, \ldots, n-1$, and

$$\begin{aligned} \|y_0 - a\| &< \varepsilon, \\ \|\Psi(y_i, t_i) - y_{i+1}\| &< \varepsilon, \quad i = 0, \ldots, n-1, \\ y_n &= b. \end{aligned}$$

Let $S$ be a compact and invariant set $S$ for $\Psi$. If for every $\varepsilon > 0$, $T > 0$ and every $a, b \in S$, there is an $(\varepsilon, T)$-pseudo orbit from $a$ to $b$, then the set $S$ is said to be *Internally Chain Transitive* (ICT). We shall say that a random

process $v(t)$ on $\mathbb{R}_+$, who is valued in $\mathbb{R}^N$, is an almost sure asymptotic pseudo trajectory [3, 4] for the differential inclusion (3) if

$$\sup_{s\in[0,T]} \|v(t+s) - \Psi(\mathrm{proj}_{\mathrm{cl}(\mathrm{dom}(\mathsf{A}))}(v(t)), s)\| \xrightarrow[t\to\infty]{} 0 \quad \text{a.s.}$$

for any $T > 0$. We note that in the APT definition provided in [3, 4], no projection is considered because the flow is defined in these references on the whole space. Projecting on $\mathrm{cl}(\mathrm{dom}(\mathsf{A}))$ here does not alter the conclusions. Let $L(v) := \bigcap_{t\geq 0} \mathrm{cl}(v([t,\infty[))$ be the *limit set* of the trajectory $v(t)$, *i.e.*, the set of the limits of the convergent subsequences $v(t_k)$ as $t_k \to \infty$. It is important to note that if $\{v(t)\}_{t\in\mathbb{R}_+}$ is bounded a.s., and if $v$ is an almost sure APT for (3), then with probability one, the compact set $L(v)$ is ICT for the semiflow $\Psi$ [3].

The authors of [8] establish a useful property of asymptotic pseudo trajectories pertaining to the asymptotic behavior of their empirical measures. We now consider that $v : \Omega \times \mathbb{R}_+ \to \mathbb{R}^N$ is a random process on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ equipped with a filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_+}$. As we know, $v$ is said to be *progressively measurable* if for each $t \geq 0$, the restriction to $\Omega \times [0,t]$ of $v$ is $\mathscr{F}_t \otimes \mathscr{B}([0,t])$-measurable, where $\mathscr{B}([0,t])$ is the Borel field over $[0,t]$. For $t \geq 0$, the *empirical measure* $\nu_t(\omega,\cdot)$ of $v$ is then the random probability measure, defined by the identity

$$\int f(x)\,\nu_t(\omega,dx) = \frac{1}{t}\int_0^t f(v(\omega,s))\,ds,$$

for any measurable function $f : \mathbb{R}^N \to \mathbb{R}_+$. We also note that a probability measure $\nu$ on $\mathbb{R}^N$ is said to be *invariant* for the semiflow $\Psi$ if

$$\int f(x)\,\nu(dx) = \int f(\Phi(x,t))\,\nu(dx)$$

for any $t \geq 0$ and any measurable function $f : \mathbb{R}^N \to \mathbb{R}_+$.

Now, if $v$ is progressively measurable and if it is an almost sure APT for the semiflow $\Psi$, then on a probability one set, all of the accumulation points of the set $\{\nu_t(\omega,\cdot)\}_{t\geq 0}$ for the weak convergence of probability measures are invariant measures for $\Psi$ [8, Th. 1]. [1]

# 3 Results

## 3.1 Algorithm Description and Main Results

Let $B : \Xi \to \mathcal{M}$ be a mapping such that, similarly to the mapping $A$ introduced in Section 2.3, the function $\xi \mapsto \mathrm{gr}(B(\xi,\cdot))$ is measurable. Moreover,

[1] The result is stated in [8] when $v$ is a so-called *weak APT*. It turns out that any almost sure APT is a weak APT by Lévy's conditional form of Borel-Cantelli's lemma.

we assume throughout the paper that $\mathrm{dom}(B(\xi,\cdot)) = \mathbb{R}^N$ for almost every $\xi \in \Xi$. We also assume that for every $x \in \mathbb{R}^N$, $B(\cdot, x)$ is integrable, and we set $\mathcal{B}(x) := \int B(\xi, x)\mu(d\xi)$. Note that $\mathrm{dom}\,\mathcal{B} = \mathbb{R}^N$. Let $(u_n)_{n\in\mathbb{N}^*}$ be an iid sequence of random variables from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\Xi, \mathscr{T})$ having the distribution $\mu$. Starting with some arbitrary $x_0 \in \mathbb{R}^N$, our purpose is to study the behavior of the iterates

$$x_{n+1} = J_{\gamma_{n+1}}(u_{n+1}, x_n - \gamma_{n+1} b(u_{n+1}, x_n)), \qquad (n \in \mathbb{N}), \tag{4}$$

where the positive sequence $(\gamma_n)_{n\in\mathbb{N}^*}$ belongs to $\ell^2 \setminus \ell^1$, and where $b$ is a measurable map on $(\Xi \times \mathbb{R}^N, \mathscr{T} \otimes \mathscr{B}(\mathbb{R}^N)) \to (\mathbb{R}^N, \mathscr{B}(\mathbb{R}^N))$ such that for every $x \in \mathbb{R}^N$, $b(\,.\,, x) \in \mathcal{S}^1_{B(\,.\,,x)}$ (2). A possible choice for $b$ is $b(\xi, x) = B_0(\xi, x)$, which is $\mathscr{T} \otimes \mathscr{B}(\mathbb{R}^N)$–measurable, as the limit as $\gamma \downarrow 0$ of $B_\gamma(\xi, x)$. We define the affine interpolated process as

$$x(t) := x_n + \frac{x_{n+1} - x_n}{\gamma_{n+1}}(t - \tau_n) \tag{5}$$

for every $t \in [\tau_n, \tau_{n+1}[$, where $\tau_n = \sum_1^n \gamma_k$. Consider the differential inclusion

$$\begin{cases} \dot{z}(t) \in -(\mathcal{A} + \mathcal{B})(z(t)), & \forall t \in \mathbb{R}_+ \text{ a.e.}, \\ z(0) = z_0\,. \end{cases} \tag{6}$$

If $\mathcal{A}+\mathcal{B}$ is maximal, then for any $z_0 \in \mathcal{D}$, (6) has a unique solution, in which case, $\Phi : \mathrm{cl}(\mathcal{D}) \times \mathbb{R}_+ \to \mathrm{cl}(\mathcal{D})$ will represent the semiflow associated to (6).

Before stating our main result, we need to make a preliminary remark. A point $x_\star$ is an element of $\mathcal{Z} = Z(\mathcal{A}+\mathcal{B})$ if and only if there exists $\varphi \in \mathcal{S}^1_{A(\cdot, x_\star)}$ and $\psi \in \mathcal{S}^1_{B(\cdot, x_\star)}$ such that $\int \varphi d\mu + \int \psi d\mu = 0$. We will refer to a couple $(\varphi, \psi)$ of this type as a *representation* of the zero $x_\star$. Moreover, in Theorem 3.1 below, we shall assume that there exists such a zero $x_\star$ for which the above functions $\varphi$ and $\psi$ can be chosen in $\mathcal{L}^{2p}(\Xi, \mathscr{T}, \mu; \mathbb{R}^N)$, where $p \geq 1$ is some integer possibly strictly larger than one. We thus introduce the set of $2p$-integrable representations

$$\mathcal{R}_{2p}(x_\star) = \left\{ (\varphi, \psi) \in \mathcal{S}^{2p}_{A(\cdot, x_\star)} \times \mathcal{S}^{2p}_{B(\cdot, x_\star)} \;:\; \int \varphi d\mu + \int \psi d\mu = 0 \right\}.$$

We let $\Pi(\xi, .)$ be the projection operator onto $\mathrm{cl}(D(\xi))$, and $d(\xi, \cdot)$ (respectively $\boldsymbol{d}(\cdot)$) be the distance function to $D(\xi)$ (respectively to $\mathcal{D}$).

**Theorem 3.1.** *Assume the following facts:*

1. *The monotone operator $\mathcal{A}$ is maximal.*

2. *There exists an integer $p \geq 1$ and a point $x_\star \in \mathcal{Z}$ such that $\mathcal{R}_{2p}(x_\star) \neq \emptyset$.*

3. *For any compact set $K$ of $\mathbb{R}^N$, there exists $\varepsilon \in ]0,1]$ such that*

$$\sup_{x \in K \cap \mathcal{D}} \int \|A_0(\xi, x)\|^{1+\varepsilon} \, \mu(d\xi) < \infty.$$

*Moreover, there exists $y_0 \in \mathcal{D}$ such that*

$$\int \|A_0(\xi, y_0)\|^{1+1/\varepsilon} \, \mu(d\xi) < \infty \, .$$

4. *There exists $C > 0$ such that for all $x \in \mathbb{R}^N$,*

$$\int d(\xi, x)^2 \mu(d\xi) \geq C \boldsymbol{d}(x)^2 \, ,$$

*and furthermore, $\gamma_{n+1}/\gamma_n \to 1$.*

5. *There exists $C > 0$ such that for any $x \in \mathbb{R}^N$ and any $\gamma > 0$,*

$$\frac{1}{\gamma^4} \int \|J_\gamma(\xi, x) - \Pi(\xi, x)\|^4 \mu(d\xi) \leq C(1 + \|x\|^{2p}) \, ,$$

*where the integer $p$ is specified in 2.*

6. *There exists $M : \Xi \to \mathbb{R}_+$ such that $M^{2p}$ is $\mu$-integrable, and for all $x \in \mathbb{R}^N$, $\|b(\xi, x)\| \leq M(\xi)(1 + \|x\|)$. Moreover, there exists a constant $C > 0$ such that $\int \|b(\xi, x)\|^4 \mu(d\xi) \leq C(1 + \|x\|^{2p})$.*

*Then, the monotone operator $\mathcal{A} + \mathcal{B}$ is maximal. Moreover, with probability one, the continuous time process $x(t)$ defined by (5) is bounded and is an APT of the differential inclusion (6).*

Let us now discuss our assumptions. Sufficient conditions for the maximality of $\mathcal{A}$ are provided below in Sections 3.2 and 4.1. Assumption 2 is relatively weak and easy to check. If we set $\varepsilon = 1$, then Assumption 3 can be replaced with the stronger condition stating that for any compact set $K$ of $\mathbb{R}^N$,

$$\sup_{x \in K \cap \mathcal{D}} \int \|A_0(\xi, x)\|^2 \, \mu(d\xi) < \infty \, .$$

For more insight on the above assumption, let us compare it with the standard Robbins-Monro algorithm $y_{n+1} = y_n + \gamma_{n+1} H(y_n, \xi_{n+1})$, where $H$ is some measurable function. In order to ensure the almost-sure boundedness of $(y_n)$, it is standard to assume that $\|H(y, \xi)\| \leq M(\xi)(1 + \|y\|)$ for every $(y, \xi)$ and for some square-integrable r.v. $M(\xi)$ [18]. As far as our algorithm is concerned, a similar assumption is needed on the operator $B$, but on the other hand, no such assumption is needed on the operator $A$. Assumption 3 is weaker. Otherwise stated, when a random operator is used through its

resolvent, there is no need to require the "linear growth" condition often assumed in the stochastic approximation literature.

Assumption 4 is quite weak, and is easy to illustrate in the case where $\mu$ is a finite sum of Dirac measures. Following [19], we say that a finite collection of closed and convex subsets $(\mathcal{C}_1, \ldots, \mathcal{C}_m)$ over some Euclidean space is *linearly regular* if there exists $\kappa > 0$ such that for every $x$,

$$\max_{i=1\ldots m} d(x, \mathcal{C}_i) \geq \kappa d(x, \mathcal{C}), \quad \text{where } \mathcal{C} = \bigcap_{i=1}^{m} \mathcal{C}_i \,,$$

and where implicitely $\mathcal{C} \neq \emptyset$. Sufficient conditions for a collection of sets to satisfy the above condition can be found in [19] and the references therein. Note that this condition implies the so-called strong conical hull intersection property $N_{\mathcal{C}}(x) = \sum_{i=1}^{m} N_{\mathcal{C}_i}(x)$ for every $x \in \mathcal{C}$, where $N_{\mathcal{C}}(x)$ is, as we recall, the normal cone to $\mathcal{C}$ at the point $x$.

Let us finally discuss Assumption 5. As $\gamma \to 0$, it is known that $J_\gamma(\xi, x)$ converges to $\Pi(\xi, x)$ for every $(\xi, x)$. Moreover, Assumption 5 provides a control on the convergence rate. The fourth moment of $\|J_\gamma(\xi, x) - \Pi(\xi, x)\|$ is assumed to vanish at the rate $\gamma^4$ with a multiplicative factor of the order $\|x\|^{2p}$. The integer $p$ can potentially be as large as needed, provided that one is able to find a zero $x_\star$ satisfying Assumption 2. In the special case where $A(\xi, \,.\,)$ coincides with the subdifferential of the convex function $f(\xi, \,.\,)$, Assumption 5 holds under the sufficient condition that for almost every $\xi$ and for every $x \in D(\xi)$,

$$\|\partial_x f_0(\xi, x)\| \leq M'(\xi)(1 + \|x\|^{p/2}) \,, \tag{7}$$

where $\partial_x f_0(\xi, x)$ is the smallest norm element of the subdifferential of $f(\xi, \,.\,)$ at point $x$, and where $M'(\xi)$ is a positive r.v. with a finite fourth moment. Indeed, in this case, the resolvent $J_\gamma(\xi, x)$ coincides with $\text{prox}_{\gamma f(\xi, \,.\,)}(x)$, and by [7],

$$\frac{1}{\gamma}\|J_\gamma(\xi, x) - \Pi(\xi, x)\| \leq 2\|\partial f_0(\xi, \Pi(\xi, x))\| \,.$$

As a consequence, Assumption 5 stems from (7) and the non-expansiveness of $\Pi(\xi, \,.\,)$.

The results of Theorem 3.1 can first be used to study the convergence of the sequence $(\bar{x}_n)$ of empirical means, defined by

$$\bar{x}_n := \frac{\sum_{k=1}^{n} \gamma_k x_k}{\sum_{k=1}^{n} \gamma_k} \,.$$

**Corollary 3.1.** *Let the assumptions in the statement of Theorem 3.1 hold true. Assume that for any $x_\star \in \mathcal{Z}$, the set $\mathcal{R}_2(x_\star)$ is not empty. Then, for any initial value $x_0$, the sequence $(\bar{x}_n)$ of empirical means converges almost surely as $n \to \infty$ to a random variable $U$, whose support lies in $\mathcal{Z}$.*

Let us now consider the issue of the convergence of the sequence $(x_n)$ to a point of $\mathcal{Z}$. Note that the conditions of Theorem 3.1 are generally insufficient to ensure that $x_n$ converges. A counterexample is obtained by setting $N = 2$ and taking $\mathcal{A}$ as a $\pi/2$-rotation matrix, $\mathcal{B} = 0$ [20, Sec. 6]. However, the statement will be proved valid when $\mathcal{A} + \mathcal{B}$ is assumed demipositive. We start by listing some known verifiable conditions ensuring that the maximal monotone operator $\mathcal{A} + \mathcal{B}$ is demipositive:

1. $\mathcal{A} + \mathcal{B} = \partial G$, where $G \in \Gamma_0$ has a minimum.
2. $\mathcal{A} + \mathcal{B} = I - T$, where $T$ is a non-expansive mapping having a fixed point.
3. The interior of $\mathcal{Z}$ is not empty.
4. $\mathcal{Z} \neq \emptyset$ and $\mathcal{A} + \mathcal{B}$ is 3-monotone, *i.e.*, for every triple $(x_i, y_i) \in \mathcal{A} + \mathcal{B}$ for $i = 1, 2, 3$, it holds that $\sum_{i=1}^{3} \langle y_i, x_i - x_{i-1} \rangle \geq 0$ by setting $x_0 = x_3$.
5. $\mathcal{A} + \mathcal{B}$ is strongly monotone, *i.e.*, $\langle x_1 - x_2, y_1 - y_2 \rangle \geq \alpha \|x_1 - x_2\|^2$ for some $\alpha > 0$ and for all $(x_1, y_1)$ and $(x_2, y_2)$ in $\mathcal{A} + \mathcal{B}$.
6. $\mathcal{Z} \neq \emptyset$ and $\mathcal{A} + \mathcal{B}$ is cocoercive, *i.e.*, $\langle x_1 - x_2, y_1 - y_2 \rangle \geq \alpha \|y_1 - y_2\|^2$ for some $\alpha > 0$ and for all $(x_1, y_1)$ and $(x_2, y_2)$ in $\mathcal{A} + \mathcal{B}$.

The above conditions can be found in [20]. Specifically, conditions 1–3 can be found in [9], while Condition 4 can be found in [21]. Conditions 5 and 6 can be easily verified to lead to the demipositivity of $\mathcal{A} + \mathcal{B}$. Condition 1 is further discussed in Section 4.1 below. Condition 2 is satisfied if $\mathcal{Z} \neq \emptyset$ and if for any $\xi$, the operator $I - (A + B)(\xi, \cdot)$ is a non-expansive mapping. Condition 4 is satisfied if $\mathcal{Z} \neq \emptyset$ and if all the operators $(A + B)(\xi, \cdot)$ are 3-monotone. The last two conditions are most often easily verifiable.
We now have:

**Corollary 3.2.** *Let the assumptions in the statement of Theorem 3.1 hold true. Assume in addition that the operator $\mathcal{A} + \mathcal{B}$ is demipositive, and that for any $x_\star \in \mathcal{Z}$, the set $\mathcal{R}_2(x_\star)$ is not empty. Then, for any initial value $x_0$, there exists a random variable $U$, supported by $\mathcal{Z}$, such that $x_n \to U$ almost surely as $n \to \infty$.*

We now address the important problem of the maximality of $\mathcal{A}$.

## 3.2 Maximality of $\mathcal{A}$

By extending a well-known result on the maximality of the sum of two maximal monotone operators, it is obvious that $\mathcal{A}$ is maximal in the case where $\mu$ is a finite sum of Dirac measures and where the interior of $\mathcal{D}$ is not empty [10, 11]. For more general measures $\mu$, we have the following result.

**Proposition 3.1.** *Assume the following:*

1. *The interior of $\mathcal{D}$ is not empty, and there exists a closed ball in $\mathcal{D}$ such that $\|A_0(\xi, x)\| \le M(\xi)$ for any $x$ in this ball, and such that $M(\xi)$ is $\mu$-integrable.*

2. *For any compact set $K$ of $\mathbb{R}^N$, there exists $\varepsilon > 0$ such that*

$$\sup_{x \in K \cap \mathcal{D}} \int \|A_0(\xi, x)\|^{1+\varepsilon} \mu(d\xi) < \infty.$$

*Moreover, there exists $y_0 \in \mathcal{D}$ such that*

$$\int \|A_0(\xi, y_0)\|^{1+1/\varepsilon} \mu(d\xi) < \infty\,.$$

3. *There exists $C > 0$ such that for any $x \in \mathbb{R}^N$,*

$$\int d(\xi, x) \mu(d\xi) \ge C \boldsymbol{d}(x).$$

4. *$\int \|J_\gamma(\xi, x) - \Pi(\xi, x)\| \mu(d\xi) \le \gamma C(x)$, where $C(x)$ is bounded on compact sets of $\mathbb{R}^N$.*

*Then, the monotone operator $\mathcal{A}$ is maximal.*

# 4 Application to Convex Optimization

We start this section by briefly reproducing some known results related to the case where $A(\xi, \cdot)$ is the subdifferential of a proper, closed and convex function $g(\xi, \cdot)$.

## 4.1 Known Facts About the Aumann Integral of Subdifferentials

A function $g : \Xi \times \mathbb{R}^N \to ]-\infty, \infty]$ is called a *normal integrand* [22] if the set-valued mapping $\xi \mapsto \operatorname{epi} g(\xi, \cdot)$ is closed-valued and measurable. Let us assume in addition that $g(\xi, \cdot)$ is convex and proper for every $\xi$.

Consider the case where $A(\xi, \cdot) = \partial g(\xi, \cdot)$. The mean operator $\mathcal{A}$ is given by[2] $\mathcal{A}(x) = \int \partial g(\xi, x) \mu(d\xi)$. Under some general conditions stated in [24], the integral and the subdifferential can be exchanged in this expression. In

[2] By [14, 23], the mapping $A : \Xi \to \mathcal{M}$, defined as $A(\xi, \cdot) = \partial g(\xi, \cdot)$, is measurable in the sense of Section 2.3.

this case, $\mathcal{A}(x) = \partial G(x)$, where $G(x) = \int g(\xi, x)\,\mu(d\xi)$. This integral is defined as the sum

$$\int_{\{\xi\,:\,g(\xi,x)\in\mathbb{R}_+\}} g(\xi, x)\,\mu(d\xi) + \int_{\{\xi\,:\,g(\xi,x)\in]-\infty,0[\}} g(\xi, x)\,\mu(d\xi) + I(x)\,,$$

where

$$I(x) = \begin{cases} +\infty, & \text{if } \mu(\{\xi : g(\xi, x) = \infty\}) > 0, \\ 0, & \text{otherwise}\,, \end{cases}$$

and where the convention $(+\infty) + (-\infty) = +\infty$ is used. The function $G$ is a lower semi continuous and convex function if $G(x) > -\infty$ for all $x$ [24]. Assuming in addition that $G$ is proper, the identity $\mathcal{A} = \partial G$ ensures that the operator $\mathcal{A}$ is monotone, maximal, and demipositive, and that the zeros of $\mathcal{A}$ are the minimizers of $G$.

### 4.2 A Constrained Optimization Problem

Let $(\mathsf{X}, \mathscr{X}, \nu)$ be a probability space. Let the functions $f : \mathsf{X} \times \mathbb{R}^N \to ]-\infty, \infty[$ and $g : \mathsf{X} \times \mathbb{R}^N \to ]-\infty, \infty[$ be normal convex integrands. Here we assume that $g$ is finite everywhere to simplify the presentation. However we note that the results can be extended to the case where $g$ is allowed to take the value $+\infty$. Recall the optimization problem

$$\min_{x\in\mathcal{C}} F(x) + G(x), \quad \mathcal{C} = \bigcap_{i=1}^{m} \mathcal{C}_i, \tag{8}$$

where $F(x) = \int f(\eta, x)\nu(d\eta)$, $G(x) = \int g(\eta, x)\nu(d\eta)$ and $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are closed and convex sets. Consider a measurable function $\tilde{\nabla} f : \mathsf{X} \times \mathbb{R}^N \to \mathbb{R}$ such that for every $\eta \in \mathsf{X}$ and $x \in \mathbb{R}^N$, $\tilde{\nabla} f(\eta, x)$ is a subgradient of $f(\eta, \,.\,)$ at $x$. Let $(v_n)_n$ be an iid sequence on $\mathsf{X}$ with probability distribution $\nu$. Finally, let $(I_n)$ be an iid sequence on $\{0, 1, \ldots, m\}$ with distribution $\alpha_i = \mathbb{P}(I_1 = i) > 0$ for every $i$. We consider the iterates

$$x_{n+1} = \begin{cases} \operatorname{prox}_{\alpha_0^{-1}\gamma_{n+1} g(v_{n+1},\cdot)}(x_n - \gamma_{n+1}\tilde{\nabla} f(v_{n+1}, x_n)), & \text{if } I_{n+1} = 0, \\ \operatorname{proj}_{\mathcal{C}_{I_{n+1}}}(x_n - \gamma_{n+1}\tilde{\nabla} f(v_{n+1}, x_n)), & \text{otherwise.} \end{cases} \tag{9}$$

We recall that $\partial g_0(\eta, x)$ is the least norm element of the subdifferential of $g(\eta, \,.\,)$ at $x$. Given $H \subset \mathbb{R}^N$, we use the notation $|H| = \sup\{\|v\| : v \in H\}$.

**Corollary 4.1.** *We assume the following. Let $p \geq 1$ be an integer.*

1. *For every $x \in \mathbb{R}^N$, $\int |f(\eta, x)|\nu(d\eta) + \int |g(\eta, x)|\nu(d\eta) < \infty$.*

2. *For any solution $x_\star$ to Problem (8), there exists a measurable function $M_\star : \mathsf{X} \to \mathbb{R}_+$ such that $\int M_\star(\eta)^2\nu(d\eta) < \infty$, and for all $\eta \in \mathsf{X}$,*

$$|\partial f(\eta, x_\star)| + |\partial g(\eta, x_\star)| \leq M_\star(\eta)\,.$$

*Moreover, there exists a solution $x_\star$ for which $\int M_\star(\eta)^{2p}\nu(d\eta) < \infty$.*

3. *For any compact set $K$ of $\mathbb{R}^N$, there exists $\varepsilon \in ]0, 1]$ such that*

$$\sup_{x \in K} \mathbb{E}\|\partial g_0(\Theta, x)\|^{1+\varepsilon} < \infty .$$

*Moreover, there exists $y_0 \in \mathcal{C}$ such that $\mathbb{E}\|\partial g_0(\Theta, y_0)\|^{1+1/\varepsilon} < \infty$.*

4. *The closed and convex sets $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are linearly regular,* i.e.,

$$\exists \kappa > 0, \forall x \in \mathbb{R}^N, \max_{i=1,\ldots,m} \operatorname{dist}(x, \mathcal{C}_i) \geq \kappa \operatorname{dist}(x, \mathcal{C}) ,$$

*where $\operatorname{dist}(x, S)$ denotes the distance of the point $x$ to the set $S$. Moreover, $\gamma_n/\gamma_{n+1} \to 1$.*

5. *There exists $M : \mathsf{X} \to \mathbb{R}$ such that $\int M(\eta)^{2p} \nu(d\eta) < \infty$, and*

$$\forall (\eta, x) \in \mathsf{X} \times \mathbb{R}^N, \ \|\tilde{\nabla} f(\eta, x)\| \leq M(\eta)(1 + \|x\|) .$$

6. *There exists $c > 0$ such that $\forall x \in \mathbb{R}^N$, $\int \|\tilde{\nabla} f(\eta, x)\|^4 \nu(d\eta) \leq c(1 + \|x\|^{2p})$.*

*Then, the sequence $(x_n)$ given by (9) converges almost surely to a solution to Problem (8).*

# 5 Related Works

The problem of minimizing an objective function in a noisy environment has brought forth a very rich body of literature in the field of stochastic approximation [17, 25]. In the framework of this paper, most of this literature examines the evolution of the projected stochastic gradient or subgradient algorithm, where the projection is made on a fixed constraining set.

In the case where the constraining set has a complicated structure, an incremental minimization algorithm with random constraint updates has been proposed in [26], where a deterministic convex function $f$ is minimized on a finite intersection of closed and convex constraining sets. The algorithm developed in [26] consists of a subgradient step over the objective $f$ followed by an update step towards a randomly chosen constraining set. Using the same principle, a distributed algorithm involving an additional consensus step has been proposed in [27]. Random iterations involving proximal and subgradient operators were considered in [28] and in [29]. In [29], the functions $g(\xi, .)$ are supposed to have a full domain, to satisfy $\|g(\xi, x) - g(\xi, y)\| \leq L(\|x - y\| + 1)$ for some constant $L$ which does not depend on $\xi$ and, finally, are such that $\int \|g(\xi, x)\|^2 \mu(d\xi) \leq L(1 + \|x\|^2)$. In the present paper, such conditions are not needed.

The algorithm (4) can also be used to solve a variational inequality problem. Let $\mathcal{C} = \cap_{i=1}^m \mathcal{C}_i$ where $\mathcal{C}_1, \ldots, \mathcal{C}_m$ are closed and convex sets in

$\mathbb{R}^N$. Consider the problem of finding $x_\star \in \mathcal{C}$ that solves the variational inequality

$$\forall x \in \mathcal{C},\ \langle F(x_\star), x - x_\star \rangle \geq 0\,,$$

where $F : \mathbb{R}^N \to \mathbb{R}^N$ is a monotone single-valued operator on $\mathbb{R}^N$ [30, 31]. Since the projection on $\mathcal{C}$ is difficult, one can use the simple stochastic algorithm $x_{n+1} = \mathrm{proj}_{\mathcal{C}_{u_{n+1}}}(x_n - \gamma_{n+1} F(x_n))$, where the random variables $u_n$ are distributed on the set $\{1, \ldots, m\}$. The variant where $F$ is itself an expectation can also be considered *i.e.*, $F(x) = \int f(\xi, x)\mu(d\xi)$. The work [30] addresses this context. In [30], it is assumed that $F$ is strongly monotone and that the stochastic Lipschitz property $\int \|f(\xi, x) - f(\xi, y)\|^2 \mu(d\xi) \leq C\|x - y\|^2$ holds, where $C$ is a positive constant. In our work, the strong monotonicity of $F$ is not needed, and the Lipschitz property is essentially replaced with the condition $\|\tilde{\nabla} f(\xi, x)\| \leq M(\xi)(1 + \|x\|)$, where $\tilde{\nabla} f(\xi, x)$ is a subgradient of $f(\xi, \cdot)$ at $x$ (for instance, the least norm one), and $M(\xi)$ satisfies a moment condition.

In the same vein as our paper, [32] considered a collection $\{A(i, \cdot)\}_{i=1}^N$ of $N$ maximal monotone operators, and studied the iterations

$$y_{n+1} \in A(\sigma_{n+1}(1), x_n)\,,\ x_{n+1} = \prod_{i=2}^{N} (I + \gamma_{n+1} A(\sigma_{n+1}(i), \cdot))^{-1}(x_n - \gamma_{n+1} y_{n+1})\,,$$

where $(\gamma_n) \in \ell^2 \setminus \ell^1$, and where $(\sigma_n)$ is a sequence of permutations of the set $\{1, \ldots, N\}$. The convergence of $(\bar{x}_n)$ to a zero of $\sum A(i, \cdot)$ is established in [32]. In the recent paper [33], a relaxed version of Algorithm (1) is considered, where $\mathsf{B}$ is cocoercive and where its output, as well as the output of the resolvent of $\mathsf{A}$, are subjected to random errors. The convergence of the iterates to a zero of $\mathsf{A} + \mathsf{B}$ is established under summability assumptions on these errors.

Regarding the convergence rate analysis, let us mention [34, 35] which investigate the performance of the algorithm $x_{n+1} = \mathrm{prox}_{\gamma_{n+1} g}(x_n - \gamma_{n+1} H_{n+1})$, where $H_{n+1}$ is a noisy estimate of the gradient $\nabla f(x_n)$. The same algorithm is addressed in [36], where the proximity operator is replaced by the resolvent of a fixed maximal monotone operator, and $H_{n+1}$ is replaced by a noisy version of a (single-valued) cocoercive operator evaluated at $x_n$. The paper [37] addresses the statistical analysis of the empirical means of the estimates obtained from the random proximal point algorithm.

This paper follows the line of thought of the recent paper [7], who studies the behavior of the random iterates $x_{n+1} = J_{n+1}(u_{n+1}, x_n)$ in a Hilbert space, and establishes the convergence of the empirical means $\bar{x}_n$ towards a zero of the mean operator $\mathcal{A}(x) = \int A(\xi, x)\,\mu(d\xi)$. In the present paper, the proximal point algorithm is replaced with the more general forward-backward algorithm. Thanks to the dynamic approach developed here, the convergences of both $(\bar{x}_n)$ and $(x_n)$ are studied.

Finally, it is worth noting that apart from the APT of Benaïm and Hirsch [3], many authors have introduced alternative concepts to analyze the asymptotic behavior of perturbed solutions to evolution systems. An important one is the notion of *almost-orbit* of [38, 39], and [40], which has been shown to be useful to analyze certain perturbed solution to differential inclusions of the form (3). The almost-orbit property is however more demanding than the APT property, and is in general harder to verify, although it can lead to finer convergence results. Fortunately, the concept of APT has been proven sufficient here to guarantee that the interpolated process $x(t)$ almost surely inherits both the ergodic and non-ergodic convergence properties of the orbits of $\Phi$.

# 6 Proofs

Let us start with the proof of Proposition 3.1 because it contains many elements of the proof of the main theorem.

## 6.1 Proof of Proposition 3.1

We recall that for any $\xi \in \Xi$ and any $\gamma > 0$, the Yosida approximation $A_\gamma(\xi, \cdot)$ is a single-valued $\gamma^{-1}$-Lipschitz monotone operator defined on $\mathbb{R}^N$. As a consequence, the operator $\mathcal{A}^\gamma : \mathbb{R}^N \to \mathbb{R}^N$, given by $\mathcal{A}^\gamma(x) = \int A_\gamma(\xi, x)\mu(d\xi)$, is a single-valued, continuous, and monotone operator defined on $\mathbb{R}^N$. As such, $\mathcal{A}^\gamma$ is maximal [10, Prop. 2.4]. Thus, given any $y \in \mathbb{R}^N$, there exists $x^\gamma \in \mathbb{R}^N$ such that $y = x^\gamma + \mathcal{A}^\gamma(x^\gamma)$. We shall find a sequence $\gamma_n \to 0$ such that $x^{\gamma_n} \to x^\star \in \mathcal{D}$ with $y - x^\star \in \mathcal{A}x^\star$. The maximality of $\mathcal{A}$ then follows by Minty's theorem [10].
Let $z_0$ and $\rho$ be respectively the centre and the radius of the ball referred to in Assumption 1, and set

$$u(\xi) = z_0 + \rho \frac{A_\gamma(\xi, x^\gamma)}{\|A_\gamma(\xi, x^\gamma)\|} \in \mathcal{D}\,,$$

where the convention $0/0 = 0$ is used. By the monotonicity of $A_\gamma(\xi, \cdot)$,

$$0 \le \int \langle x^\gamma - u(\xi), A_\gamma(\xi, x^\gamma) - A_\gamma(\xi, u(\xi))\rangle \, \mu(d\xi).$$

Writing $C = \int M(\xi)\mu(d\xi) < \infty$ (see Assumption 1), we obtain

$$\begin{aligned}
\int \langle x^\gamma, A_\gamma(\xi, x^\gamma)\rangle \,\mu(d\xi) &= \langle x^\gamma, y\rangle - \|x^\gamma\|^2, \\
\int \langle -u(\xi), A_\gamma(\xi, x^\gamma)\rangle \,\mu(d\xi) &= \langle z_0, x^\gamma - y\rangle - \rho \int \|A_\gamma(\xi, x^\gamma)\| \,\mu(d\xi), \\
\int |\langle x^\gamma, A_\gamma(\xi, u(\xi))\rangle| \,\mu(d\xi) &\le \|x^\gamma\| \int \|A_0(\xi, u(\xi)\| \,\mu(d\xi) \le C\|x^\gamma\|, \\
\int |\langle u(\xi), A_\gamma(\xi, u(\xi))\rangle| \,\mu(d\xi) &\le C(\|z_0\| + \rho).
\end{aligned}$$

Therefore,

$$\rho \int \|A_\gamma(\xi, x^\gamma)\| \,\mu(d\xi) + \|x^\gamma\|^2 \le \|x^\gamma\|(\|y\| + \|z_0\| + C) + C(\|z_0\| + \rho) + \|z_0\| \, \|y\| \,.$$

This shows that the sets $\{\|x^\gamma\|\}$ and $\{\int \|A_\gamma(\xi, x^\gamma)\| \,\mu(d\xi)\}$ are both bounded. Writing $A_\gamma(\xi, x^\gamma) = \gamma^{-1}(\Pi(\xi, x^\gamma) - J_\gamma(\xi, x^\gamma)) + \gamma^{-1}(x^\gamma - \Pi(\xi, x^\gamma))$, and using Assumption 4, we obtain that the set $\{\gamma^{-1} \int \|x^\gamma - \Pi(\xi, x^\gamma)\| \mu(d\xi)\}$ is bounded. By Assumption 3, $\{\boldsymbol{d}(x^\gamma)/\gamma\}$ is bounded. Given $x^\gamma$, let us choose $\tilde{x}^\gamma \in \mathcal{D}$ such that $\|x^\gamma - \tilde{x}^\gamma\| \le 2\boldsymbol{d}(x^\gamma)$. By the boundedness of $\{\|x^\gamma\|\}$, there exists a compact set $K \subset \mathbb{R}^N$ such that $\tilde{x}^\gamma \in K$. Associating a positive number $\varepsilon$ to $K$ as in Assumption 2, we obtain

$$\begin{aligned}
&\int \|A_\gamma(\xi, x^\gamma)\|^{1+\varepsilon} \,\mu(d\xi) \\
&\qquad \le 2^\varepsilon \int \Big( \|A_\gamma(\xi, \tilde{x}^\gamma)\|^{1+\varepsilon} + \|A_\gamma(\xi, x^\gamma) - A_\gamma(\xi, \tilde{x}^\gamma)\|^{1+\varepsilon} \Big) \mu(d\xi) \\
&\qquad \le 2^\varepsilon \int \|A_0(\xi, \tilde{x}^\gamma)\|^{1+\varepsilon} \,\mu(d\xi) + 2^{1+2\varepsilon} \Big| \frac{\boldsymbol{d}(x^\gamma)}{\gamma} \Big|^{1+\varepsilon},
\end{aligned}$$

which is bounded by a constant independent of $\gamma$ thanks to Assumption 2. Thus, the family of $\Xi \to \mathbb{R}^N$ functions $\{A_\gamma(\xi, x^\gamma)\}$ is bounded in the Banach space $\mathcal{L}^{1+\varepsilon}(\Xi, \mathscr{T}, \mu; \mathbb{R}^N)$.

Let us take a sequence $(\gamma_n, x^{\gamma_n})$ converging to $(0, x^\star)$. Let us extract a subsequence (still denoted as $(n)$) from the sequence of indices $(n)$, in such a way that $(A_{\gamma_n}(\xi, x^{\gamma_n}))_n$ converges weakly in $\mathcal{L}^{1+\varepsilon}$ towards a function $f(\xi)$. By Mazur's theorem, there exists a function $J : \mathbb{N} \to \mathbb{N}$ and a sequence of sets of weights $(\{\alpha_{k,n}, k = n \ldots, J(n) \,:\, \alpha_{k,n} \ge 0, \sum_{k=n}^{J(n)} \alpha_{k,n} = 1\})_n$ such that the sequence of functions $(g_n(\xi) = \sum_{k=n}^{J(n)} \alpha_{k,n} A_{\gamma_k}(\xi, x^{\gamma_k}))$ converges strongly to $f$ in $\mathcal{L}^{1+\varepsilon}$. Taking a further subsequence, we obtain the $\mu$-almost everywhere convergence of $(g_n)$ to $f$.

Observe that $x^\star \in \mathrm{cl}(\mathcal{D})$ since $\boldsymbol{d}(x^{\gamma_n}) \to 0$. Choose a sequence $(z_n)$ in $\mathcal{D}$ that converges to $x^\star$, and for each $n$, let $T_n = \{\xi \in \Xi \,:\, z_n \in D(\xi)\}$. Then, on the probability one set $T = \cap_n T_n$, it holds that $x^\star \in \mathrm{cl}(D(\xi))$. On the

intersection of $T$ and the set where $g_n \to f$, set $\eta_n(\xi) = J_{\gamma_n}(\xi, x^{\gamma_n}) - x^\star$, and write

$$\|\eta_n(\xi)\| \le \|J_{\gamma_n}(\xi, x^{\gamma_n}) - J_{\gamma_n}(\xi, x^\star)\| + \|J_{\gamma_n}(\xi, x^\star) - x^\star\|.$$

Since $J_{\gamma_n}(\xi, \cdot)$ is non-expansive and since $x^\star \in \mathrm{cl}(D(\xi))$, we have $\eta_n(\xi) \to_n 0$. Considering Assumption 2, we also have

$$\begin{aligned} \|\eta_n(\xi)\| &\le \|x^\star\| + \|J_{\gamma_n}(\xi, x^{\gamma_n}) - J_{\gamma_n}(\xi, y_0)\| + \|J_{\gamma_n}(\xi, y_0) - y_0\| + \|y_0\| \\ &\le \|x^\star\| + \sup_\gamma \|x^\gamma\| + 2\|y_0\| + \|A_0(\xi, y_0)\|, \end{aligned}$$

when $\gamma_n \le 1$. By Assumption 2 and the dominated convergence theorem, we obtain that $\eta_n \to 0$ in $\mathcal{L}^{1+1/\varepsilon}$. With this in mind,

$$\begin{aligned} &\int |\langle \eta_n(\xi), A_{\gamma_n}(\xi, x^{\gamma_n})\rangle| \mu(d\xi) \\ &\le \left( \int \|\eta_n(\xi)\|^{1+1/\varepsilon} \mu(d\xi) \right)^{\varepsilon/(1+\varepsilon)} \left( \int \|A_{\gamma_n}(\xi, x^{\gamma_n})\|^{1+\varepsilon} \mu(d\xi) \right)^{1/(1+\varepsilon)}, \end{aligned}$$

and the left-hand side converges to zero. Consequently, the random variable

$$e_n = \sum_{k=n}^{J(n)} \alpha_{k,n} \langle J_{\gamma_k}(\xi, x^{\gamma_k}) - x^\star, A_{\gamma_k}(\xi, x^{\gamma_k})\rangle$$

converges to zero in probability, hence in the $\mu$-almost sure sense along a subsequence. Fix $\xi$ in this new probability one set, choose arbitrarily a couple $(u, v) \in A(\xi, \cdot)$, and write

$$X_n = \sum_{k=n}^{J(n)} \langle u - J_{\gamma_k}(\xi, x^{\gamma_k}), \alpha_{k,n} v - \alpha_{k,n} A_{\gamma_k}(\xi, x^{\gamma_k})\rangle.$$

It holds by the monotonicity of $A(\xi, \cdot)$ that $X_n \ge 0$. Writing

$$X_n = \langle u - x^\star, v - g_n(\xi)\rangle + e_n - \sum_{k=n}^{J(n)} \alpha_{k,n} \langle \eta_k, v\rangle,$$

and making $n \to \infty$, we obtain that $\langle u - x^\star, v - f(\xi)\rangle \ge 0$. By the maximality of $A(\xi, \cdot)$, it holds that $(x^\star, f(\xi)) \in A(\xi, \cdot)$.
To conclude, we have

$$y = \sum_{k=n}^{J(n)} \alpha_{k,n} x^{\gamma_k} + \int g_n(\xi)\, \mu(d\xi),$$

$\sum_{k=n}^{J(n)} \alpha_{k,n} x^{\gamma_k} \to_n x^\star \in \mathcal{D}$, and $g_n \xrightarrow{\mathcal{L}^1(\mu)} f \in \mathcal{S}^1_{A(\cdot, x^\star)}$. Making $n \to \infty$, we obtain $y - x^\star = \int f(\xi)\, \mu(d\xi) \in \mathcal{A}(x^\star)$, which is the desired result. □

### 6.2 Proof of Theorem 3.1

Noting that $\operatorname{dom} \mathcal{B} = \mathbb{R}^N$ and using Assumption 6 of Theorem 3.1, one can check that the assumptions of Proposition 3.1 are satisfied for $B$. The result is that $\mathcal{B}$ is maximal. Because $\mathcal{B}$ has a full domain and $\mathcal{A}$ is maximal, $\mathcal{A}+\mathcal{B}$ is maximal by [11, Corollary 24.4]. Thus, the first assertion of Theorem 3.1 is shown, and moreover, the differential inclusion (6) admits a unique solution, and the associated semiflow $\Phi$ is well defined.

Defining $Y_\gamma(\xi, x) := A_\gamma(\xi, x - \gamma b(\xi, x))$, the iterates $x_n$ can be rewritten as

$$\begin{aligned} x_{n+1} &= x_n - \gamma_{n+1} b(u_{n+1}, x_n) - \gamma_{n+1} Y_{\gamma_{n+1}}(u_{n+1}, x_n) \\ &= x_n - \gamma_{n+1} h_{\gamma_{n+1}}(x_n) + \gamma_{n+1}\eta_{n+1}, \end{aligned}$$

where we define

$$h_\gamma(x) := \int (Y_\gamma(\xi, x) + b(\xi, x))\mu(d\xi)\,,$$

and

$$\eta_{n+1} := -Y_{\gamma_{n+1}}(u_{n+1}, x_n) + \mathbb{E}_n Y_{\gamma_{n+1}}(u_{n+1}, x_n) - b(u_{n+1}, x_n) + \mathbb{E}_n b(u_{n+1}, x_n)\,,$$

where $\mathbb{E}_n$ denotes the expectation conditionally to the sub $\sigma$-field $\sigma(u_1, \ldots, u_n)$ of $\mathscr{F}$ (we also write $\mathbb{E}_0 = \mathbb{E}$). Consider the martingale

$$M_n := \sum_{k=1}^n \gamma_k \eta_k\,,$$

and let $M(t)$ be the affine interpolated process, defined for any $n \in \mathbb{N}$ and any $t \in [\tau_n, \tau_{n+1}[$ as

$$M(t) := M_n + \eta_{n+1}(t - \tau_n) = M_n + \frac{M_{n+1} - M_n}{\gamma_{n+1}}(t - \tau_n).$$

For any $t \geq 0$, let

$$r(t) := \max\{k \geq 0 \;:\; \tau_k \leq t\}.$$

Then, for any $t \geq 0$, we obtain

$$\begin{aligned} x(\tau_n + t) - x(\tau_n) &= -\int_0^t h_{\gamma_{r(\tau_n+s)+1}}(x_{r(\tau_n+s)})\, ds + M(\tau_n + t) - M(\tau_n) \\ &= H(\tau_n + t) - H(\tau_n) + M(\tau_n + t) - M(\tau_n)\,, \end{aligned} \tag{10}$$

where $H(t) := \int_0^t h_{\gamma_{r(s)+1}}(x_{r(s)})\, ds$. The idea of the proof is to establish that on a $\mathbb{P}$-probability one set, the sequence $(x(\tau_n + \cdot))_{n\in\mathbb{N}}$ of continuous time processes is equicontinuous and bounded. The accumulation points for the

uniform convergence on a compact interval $[0,T]$ (who are guaranteed to exist by the Arzelà-Ascoli theorem) will be shown to have the form

$$z(t)-z(0)=-\lim_{n\to\infty}\int_0^t ds\int_\Xi \mu(d\xi)\,(Y_{\gamma_{r(\tau_n+s)+1}}(\xi,x_{r(\tau_n+s)})+b(\xi,x_{r(\tau_n+s)}))\,, \tag{11}$$

where the limit is taken over a subsequence. We then show that the sequence of $\Xi\times[0,T]\to\mathbb{R}^{2N}$ functions $((\xi,s)\mapsto Y_{\gamma_{r(\tau_n+s)+1}}(\xi,x_{r(\tau_n+s)}),b(\xi,x_{r(\tau_n+s)}))_n$ is bounded in the Banach space $\mathcal{L}^{1+\varepsilon}(\Xi\times[0,T],\mu\otimes\lambda)$, where $\lambda$ is the Lebesgue measure on $[0,T]$. Analyzing the accumulation points and following an approach similar to the one used in the proof of Proposition 3.1, we prove that the limit in the right-hand side of (11) coincides with

$$z(t)-z(0)=-\lim_{n\to\infty}\int_0^t ds\left(\int_\Xi f^{(a)}(\xi,s)\mu(d\xi)+\int_\Xi f^{(b)}(\xi,s)\mu(d\xi)\right)\,,$$

where for almost every $s\in[0,T]$, $f^{(a)}(\cdot,s)$ and $f^{(b)}(\cdot,s)$ are integrable selections of $A(\cdot,s)$ and $B(\cdot,s)$, respectively. This shows that $z$ satisfies the differential inclusion (6). Hence, almost surely, the accumulation points of the sequence of processes $(x(\tau_n+\cdot))_{n\in\mathbb{N}}$ are solutions to (6). Recalling that the latter defines a semiflow $\Phi:\mathrm{cl}(\mathcal{D})\times\mathbb{R}_+\to\mathrm{cl}(\mathcal{D})$, it follows that the process $x(t)$ is a.s. an APT of (6).

Throughout the proof, $C$ refers to a positive constant, that can change from line to line, but that remains independent of $n$. We use $c$, $c_1$, etc. to denote random variables on $\Omega\to\mathbb{R}_+$ that do not depend on $n$. For a fixed event $\omega\in\Omega$, these will act as constants.

**Proposition 6.1.** *Let Assumptions 2 and 6 of Theorem 3.1 hold true. Then,*

1. *The sequence $(x_n)$ is bounded almost surely and in $\mathcal{L}^2(\Omega,\mathscr{F},\mathbb{P};\mathbb{R}^N)$.*
2. $\mathbb{E}[\sum_n\gamma_n^2\int\|Y_{\gamma_n}(\xi,x_n)\|^2\mu(d\xi)]<\infty$.
3. *The sequence $(\|x_n-x_\star\|)_n$ converges almost surely.*

*Proof.* Writing $\|x_{n+1}-x_\star\|^2=\|x_n-x_\star\|^2+2\langle x_{n+1}-x_n,x_n-x_\star\rangle+\|x_{n+1}-x_n\|^2$, we obtain

$$\begin{aligned}&\|x_{n+1}-x_\star\|^2=\|x_n-x_\star\|^2-2\gamma_{n+1}\langle Y_{\gamma_{n+1}}(u_{n+1},x_n),x_n-x_\star\rangle\\&\quad-2\gamma_{n+1}\langle b(u_{n+1},x_n),x_n-x_\star\rangle+\gamma_{n+1}^2\|b(u_{n+1},x_n)+Y_{\gamma_{n+1}}(u_{n+1},x_n)\|^2.\end{aligned}$$

Thanks to Assumption 2, we can choose $\varphi\in\mathcal{S}^2_{A(\cdot,x_\star)}$ and $\psi\in\mathcal{S}^1_{B(\cdot,x_\star)}$ such that $0=\int(\varphi+\psi)d\mu$. Writing $u=u_{n+1}$, $\gamma=\gamma_{n+1}$, $Y_\gamma=Y_{\gamma_{n+1}}(u_{n+1},x_n)$,

$J_\gamma = J_{\gamma_{n+1}}(u_{n+1}, x_n - \gamma_{n+1} b(u_{n+1}, x_n))$, and $b = b(u_{n+1}, x_n)$ for conciseness, and recalling that $Y_\gamma = (x - \gamma b - J_\gamma)/\gamma$, we write

$$\begin{aligned}\langle Y_\gamma, x_n - x_\star\rangle &= \langle Y_\gamma - \varphi(u), J_\gamma - x_\star\rangle + \gamma\langle Y_\gamma - \varphi(u), Y_\gamma\rangle + \gamma\langle Y_\gamma - \varphi(u), b\rangle \\ &\quad + \langle \varphi(u), x_n - x_\star\rangle \\ &\geq \gamma\|Y_\gamma\|^2 - \gamma\langle\varphi(u), Y_\gamma\rangle + \gamma\langle Y_\gamma - \varphi(u), b\rangle + \langle\varphi(u), x_n - x_\star\rangle,\end{aligned}$$

since $Y_\gamma \in A(u, J_\gamma)$ and $A(\xi, \cdot)$ is monotone. By the monotonicity of $B(\xi, \cdot)$, we also have $\langle b, x_n - x_\star\rangle \geq \langle \psi(u), x_n - x_\star\rangle$. By expanding $\gamma^2\|b + Y_\gamma\|^2$, we obtain altogether

$$\begin{aligned}\|x_{n+1} - x_\star\|^2 &\leq \|x_n - x_\star\|^2 - \gamma^2\|Y_\gamma\|^2 + 2\gamma^2\langle\varphi(u), Y_\gamma\rangle + 2\gamma^2\langle\varphi(u), b\rangle \\ &\quad + \gamma^2\|b\|^2 - 2\gamma\langle\varphi(u) + \psi(u), x_n - x_\star\rangle \\ &\leq \|x_n - x_\star\|^2 - \gamma^2(1 - \beta^{-1})\|Y_\gamma\|^2 + \gamma^2(1 + \beta^{-1})\|b\|^2 \\ &\quad + 2\gamma^2\beta\|\varphi(u)\|^2 - 2\gamma\langle\varphi(u) + \psi(u), x_n - x_\star\rangle\,, \end{aligned} \tag{12}$$

where we used the inequality $|\langle a, b\rangle| \leq (\beta/2)\|a\|^2 + \|b\|^2/(2\beta)$, where $\beta > 0$ is arbitrary. By Assumption 6,

$$\mathbb{E}_n\|b\|^2 \leq C(1 + \|x_n\|^2) \leq 2C(1 + \|x_\star\|^2 + \|x_n - x_\star\|^2)$$

for some (other) constant $C$. Moreover $\mathbb{E}_n\langle\varphi(u) + \psi(u), x_n - x_\star\rangle = 0$. Thus,

$$\begin{aligned}\mathbb{E}_n\|x_{n+1} - x_\star\|^2 &\leq (1 + C\gamma_{n+1}^2)\|x_n - x_\star\|^2 \\ &\quad - \gamma_{n+1}^2(1 - \beta^{-1})\int \|Y_{\gamma_{n+1}}(\xi, x_n)\|^2\mu(d\xi) + C\gamma_{n+1}^2.\end{aligned}$$

Choose $\beta > 1$. Using the Robbins-Siegmund Lemma [41] along with $(\gamma_n) \in \ell^2$, the conclusion follows. □

*Remark* 1. This proposition calls for some comments. In the standard forward-backward algorithm described in the introduction of this paper, the operators $\mathsf{A}$ and $\mathsf{B}$ are both deterministic, and $\mathsf{B}$ is a single-valued operator satisfying a so-called cocoercivity property. In these conditions, the iteration (1) belongs to the class of the so-called *Krasnosel'skiĭ-Mann* iterations, provided the fixed step size $\gamma$ is chosen small enough [11]. A well known property of these iterations is that the sequence $(x_n)$ is *Fejér monotone* with respect to $Z(\mathsf{A} + \mathsf{B})$. Specifically, for all $x_\star \in Z(\mathsf{A} + \mathsf{B})$, $(\|x_n - x_\star\|)$ is decreasing. In our situation, the forward operators $B(\xi, \cdot)$ are not required to be single-valued. On the other hand, Assumptions 2 and 6 are needed along with the fact that $(\gamma_n) \in \ell^2$. Instead of the Fejér monotonicity, we obtain the weaker result given by Proposition 6.1-3.

The following lemma provides a moment control over the iterates $x_n$.

**Lemma 6.1.** *Let Assumptions 2 and 6 in the statement of Theorem 3.1 hold true. Then,* $\sup_n \mathbb{E}\|x_n\|^{2p} < \infty$.

*Proof.* We shall establish the result by recurrence over $p$. Proposition 6.1 shows that it holds for $p = 1$. Assume that it holds for $p - 1$. Using Assumption 2, choose $\varphi \in \mathcal{S}^{2p}_{A(\,.\,,x_\star)}$ and $\psi \in \mathcal{S}^{2p}_{B(\,.\,,x_\star)}$ such that $0 = \int(\varphi + \psi)d\mu$. Inequality (12) shows that for some constant $C > 0$,

$$\|x_{n+1} - x_\star\|^2 \le \|x_n - x_\star\|^2 - 2\gamma_{n+1}\langle \varphi(u_{n+1}) + \psi(u_{n+1}), x_n - x_\star\rangle \\ + C\gamma_{n+1}^2(\|\varphi(u_{n+1})\|^2 + \|b(u_{n+1}, x_n)\|^2)\,.$$

Raising both sides to the power $p$ then taking their expectations, we obtain

$$\mathbb{E}\|x_{n+1} - x_\star\|^{2p} \le \sum_{k_1+k_2+k_3=p} \frac{p!}{k_1!k_2!k_3!} C^{k_2}(-2)^{k_3}\gamma_{n+1}^{2k_2+k_3} T_n^{(k_1,k_2,k_3)}\,, \tag{13}$$

where we set for every $\vec{k} = (k_1, k_2, k_3)$,

$$T_n^{\vec{k}} = \mathbb{E}\Big[\|x_n - x_\star\|^{2k_1} \times (\|\varphi(u_{n+1})\|^2 + \|b(u_{n+1}, x_n)\|^2)^{k_2} \\ \times \langle \varphi(u_{n+1}) + \psi(u_{n+1}), x_n - x_\star\rangle^{k_3}\Big]\,.$$

We can make the following observations:

- By choosing $k_2 = k_3 = 0$, we observe that $\mathbb{E}\|x_{n+1} - x_\star\|^{2p}$ is no greater than $\mathbb{E}\|x_n - x_\star\|^{2p}$ plus some additional terms involving only smaller powers of $\|x_n - x_\star\|$.

- The term corresponding to $(k_1, k_2, k_3) = (p-1, 0, 1)$ is zero since $u_{n+1}$ and $\sigma(u_1, \ldots, u_n)$ are independent and $\mathbb{E}_n\langle \varphi(u_{n+1}) + \psi(u_{n+1}), x_n - x_\star\rangle = 0$. This implies that any term in the sum except $\mathbb{E}\|x_n - x_\star\|^{2p}$ is multiplied by $\gamma_{n+1}$, raised to a power greater than 2.

- Consider the case $(k_1, k_2, k_3) \neq (p-1, 0, 1)$ and $(k_1, k_2, k_3) \neq (p, 0, 0)$. Using Jensen's inequality and the inequality $x^k y^\ell \le x^{k+\ell} + y^{k+\ell}$ for non-negative $x, y, k$ and $\ell$, we get

$$\begin{aligned}|T_n^{\vec{k}}| &\le \mathbb{E}\Big[\|x_n - x_\star\|^{2k_1+k_3} \times (\|\varphi(u_{n+1})\|^2 + \|b(u_{n+1}, x_n)\|^2)^{k_2} \\ &\qquad\qquad \times \|\varphi(u_{n+1}) + \psi(u_{n+1})\|^{k_3}\Big] \\ &\le C\mathbb{E}\Big[\|x_n - x_\star\|^{2k_1+k_3} \times (\|\varphi(u_{n+1})\|^{2k_2} + \|b(u_{n+1}, x_n)\|^{2k_2}) \\ &\qquad\qquad \times (\|\varphi(u_{n+1})\|^{k_3} + \|\psi(u_{n+1})\|^{k_3})\Big] \\ &\le C\mathbb{E}\Big[\|x_n - x_\star\|^{2k_1+k_3}\|b(u_{n+1}, x_n)\|^{2k_2+k_3}\Big] \\ &\quad + C\mathbb{E}\Big[\|x_n - x_\star\|^{2k_1+k_3}\Big]\mathbb{E}\Big[\|\varphi(u_{n+1})\|^{2k_2+k_3} + \|\psi(u_{n+1})\|^{2k_2+k_3}\Big].\end{aligned}$$

By conditioning on $\sigma(u_1,\ldots,u_n)$ and by using Assumption 6, we get

$$\begin{aligned}&\mathbb{E}\Big[\|x_n-x_\star\|^{2k_1+k_3}\|b(u_{n+1},x_n)\|^{2k_2+k_3}\Big]\\&\le C\mathbb{E}\Big[\|x_n-x_\star\|^{2k_1+k_3}(1+\|x_n\|^{2k_2+k_3})\Big]\le C(\mathbb{E}\|x_n-x_\star\|^{2p}+1).\end{aligned}$$

Noting that $2k_1+k_3\le 2(p-1)$, we get that $\mathbb{E}\|x_n-x_\star\|^{2k_1+k_3}<C$ by the induction hypothesis. Since $2k_2+k_3\le 2p$ and since $\varphi$ and $\psi$ are $2p$-integrable selections, it follows that $|T_n^{\vec{k}}|\le C(1+\mathbb{E}\|x_n-x_\star\|^{2p})$. Note also that in the considered case, one has $2k_2+k_3\ge 2$, which implies that all terms $T_n^{\vec{k}}$ are multiplied by $\gamma_{n+1}^2$.

In conclusion, we obtain that

$$\mathbb{E}\|x_{n+1}-x_\star\|^{2p}\le\mathbb{E}(1+C\gamma_{n+1}^2)\|x_n-x_\star\|^{2p}+C\gamma_{n+1}^2$$

for some constant $C>0$. Starting from $n=0$ and iterating, we obtain that $\sup_n\mathbb{E}\|x_n-x_\star\|^{2p}<\infty$. □

We now need to control the distances to $\mathcal{D}$ of the iterates $x_n$. Let us start with an easy technical result, whose proof is left to the reader.

**Lemma 6.2.** *For any $\varepsilon>0$, there exist $C(\varepsilon)>0$ and $C'(\varepsilon)>0$ such that for any vectors $x,y\in\mathbb{R}^N$,*

$$\|x+y\|^2\le(1+\varepsilon)\|x\|^2+C(\varepsilon)\|y\|^2,\quad\text{and}\quad\|x+y\|^4\le(1+\varepsilon)\|x\|^4+C'(\varepsilon)\|y\|^4.$$

**Proposition 6.2.** *Let Assumptions 2, 4, 5, and 6 of Theorem 3.1 hold true. Then, $\boldsymbol{d}(x_n)$ tends a.s. to zero. Moreover, for every $\omega$ in a probability one set, there exists $c(\omega)>0$ and a positive sequence $(c_m(\omega))_{m\in\mathbb{N}}$ converging to zero such that for every integer $n$ and every integer $m$ such that $n\ge m$,*

$$\sum_{k=m}^{n}\frac{\boldsymbol{d}(x_k)^2}{\gamma_k}\le c_m(\omega)+c(\omega)\sum_{k=m}^{n}\gamma_k\,.$$

*Proof.* We start by writing $x_{n+1}=\Pi(u_{n+1},x_n)+\gamma_{n+1}\delta_{n+1}$, where

$$\delta_{n+1}=\frac{J_{\gamma_{n+1}}(u_{n+1},x_n-\gamma_{n+1}b(u_{n+1},x_n))-\Pi(u_{n+1},x_n)}{\gamma_{n+1}}.$$

Upon noting that $J_\gamma(\xi,\,.\,)$ is non-expansive for every $\xi$,

$$\|\delta_{n+1}\|\le\|b(u_{n+1},x_n)\|+\frac{\|J_{\gamma_{n+1}}(u_{n+1},x_n)-\Pi(u_{n+1},x_n)\|}{\gamma_{n+1}}.$$

Using Assumptions 5 and 6, we have

$$\begin{aligned}\mathbb{E}_n\|\delta_{n+1}\|^4&=4\int\|b(\xi,x_n)\|^4\mu(d\xi)+4\gamma_{n+1}^{-4}\int\|J_{\gamma_{n+1}}(\xi,x_n)-\Pi(\xi,x_n)\|^4\mu(d\xi)\\&\le C(1+\|x_n\|^{2p}).\end{aligned}$$

Therefore, by Proposition 6.1-1., there exists a non-negative $c_1(\omega)$, which is a.s. finite and satisfies $\mathbb{E}_n\|\delta_{n+1}\|^4 \le c_1(\omega)$ almost surely. By Lemma 6.1, it also holds that $\sup_n \mathbb{E}\|\delta_n\|^4 < \infty$.
Consider an arbitrary point $u \in \mathrm{cl}(\mathcal{D})$. For any $\varepsilon > 0$, by Lemma 6.2, we have

$$\|x_{n+1} - u\|^2 \le (1+\varepsilon)\|\Pi(u_{n+1}, x_n) - u\|^2 + \gamma_{n+1}^2 C\|\delta_{n+1}\|^2.$$

Since $\Pi(u_{n+1}, \cdot)$ is firmly non-expansive as the projector onto a closed and convex set, we have

$$\|\Pi(u_{n+1}, x_n) - u\|^2 \le \|x_n - u\|^2 - \|\Pi(u_{n+1}, x_n) - x_n\|^2.$$

Taking $u = \mathbf{\Pi}(x_n)$, we obtain

$$\begin{aligned}\boldsymbol{d}(x_{n+1})^2 &\le \|x_{n+1} - \mathbf{\Pi}(x_n)\|^2 \\ &\le (1+\varepsilon)(\boldsymbol{d}(x_n)^2 - d(u_{n+1}, x_n)^2) + C\gamma_{n+1}^2\|\delta_{n+1}\|^2.\end{aligned}$$

Taking the conditional expectation $\mathbb{E}_n$ at both sides of this inequality, using Assumption 4 and choosing $\varepsilon$ small enough, we obtain the inequality $\mathbb{E}_n\boldsymbol{d}^2(x_{n+1}) \le \rho\boldsymbol{d}^2(x_n) + \gamma_{n+1}^2 C\mathbb{E}_n\|\delta_{n+1}\|^2$, where $\rho \in [0,1[$. It implies that $\boldsymbol{d}^2(x_n)$ tends to zero by the Robbins-Siegmund Theorem [41]. Moreover, setting $\Delta_n = \boldsymbol{d}(x_n)^2/\gamma_n$ and using the fact that $\gamma_n/\gamma_{n+1} \to 1$, we obtain that

$$\mathbb{E}_n\Delta_{n+1} \le \rho\Delta_n + \gamma_{n+1}C\mathbb{E}_n\|\delta_{n+1}\|^2$$

for $n$ larger than some $n_0$.
By Lemma 6.2 and the firm non-expansiveness of $\Pi(u_{n+1}, \cdot)$, we also have

$$\begin{aligned}\|x_{n+1} - u\|^4 &\le (1+\varepsilon)\|\Pi(u_{n+1}, x_n) - u\|^4 + \gamma_{n+1}^4 C\|\delta_{n+1}\|^4 \\ &\le (1+\varepsilon)(\|x_n - u\|^2 - \|\Pi(u_{n+1}, x_n) - x_n\|^2)^2 + \gamma_{n+1}^4 C\|\delta_{n+1}\|^4.\end{aligned} \tag{14}$$

We also set $u = \mathbf{\Pi}(x_n)$ and apply the operator $\mathbb{E}_n$ at both sides of this inequality. By Assumption 4, we have

$$\begin{aligned}\int (\boldsymbol{d}(x)^2 - d(\xi, x)^2)^2 \mu(d\xi) &= \boldsymbol{d}(x)^4 + \int d(\xi, x)^4 \mu(d\xi) - 2\boldsymbol{d}(x)^2 \int d(\xi, x)^2 \mu(d\xi) \\ &\le \boldsymbol{d}(x)^4 - \boldsymbol{d}(x)^2 \int d(\xi, x)^2 \mu(d\xi) \le (1 - C)\boldsymbol{d}(x)^4\end{aligned}$$

since $d(\xi, x) \le \boldsymbol{d}(x)$. Integrating (14), we obtain

$$\mathbb{E}_n\boldsymbol{d}^4(x_{n+1}) \le \rho\boldsymbol{d}^4(x_n) + \gamma_{n+1}^4 C\mathbb{E}_n\|\delta_{n+1}\|^4\,,$$

where $\rho \in [0,1[$, hence $\mathbb{E}_n\Delta_{n+1}^2 \le \rho\Delta_n^2 + \gamma_n^2 C\mathbb{E}_n\|\delta_{n+1}\|^4$ for $n$ larger than some $n_0$. Taking the expectation at each side, iterating, and using the

boundedness of $(\mathbb{E}\|\delta_n\|^4)$, we obtain that $\mathbb{E}\Delta_n^2 \le C(\rho^n + \sum_{k=1}^n \gamma_k^2 \rho^{n-k})$. Therefore,

$$\sum_{n=0}^{\infty} \mathbb{E}\Delta_n^2 \le C\Big(1 + \sum_{n=0}^{\infty} \gamma_n^2\Big) < \infty.$$

Consequently, $\Delta_n \to 0$ almost surely. Moreover, the martingale

$$Y_n = \sum_{k=1}^{n} (\Delta_k - \mathbb{E}_{k-1}\Delta_k)$$

converges almost surely and in $\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R})$. Letting $D_m^n = \sum_{k=m+1}^n \Delta_k$, where $m$ and $n$ are any two integers such that $0 < m < n$, we can write

$$\begin{aligned}
D_m^n &= \sum_{k=m+1}^{n} \mathbb{E}_{k-1}\Delta_k + Y_n - Y_m \\
&\le \rho \sum_{k=m}^{n-1} (\Delta_k + C\gamma_{k+1}\mathbb{E}_k\|\delta_{k+1}\|^2) + Y_n - Y_m \\
&\le \rho\Delta_m + \rho D_m^n + \rho C\sqrt{c_1(\omega)} \sum_{k=m+1}^{n} \gamma_k + Y_n - Y_m.
\end{aligned}$$

To conclude, we have

$$D_m^n \le \frac{\rho}{1-\rho}\Delta_m + \frac{Y_n - Y_m}{1-\rho} + \frac{\rho C\sqrt{c_1(\omega)}}{1-\rho} \sum_{k=m+1}^{n} \gamma_k \,.$$

Since $\Delta_m \to 0$, and since $(Y_n(\omega))_{n\in\mathbb{N}}$ is almost surely a Cauchy sequence, we obtain the desired result. □

**Lemma 6.3.** *Let Assumptions 3 and 6 hold true. For any compact set $K$, there exists a constant $C > 0$ and $\varepsilon \in ]0, 1]$ such that for all $x \in K$ and all $\gamma > 0$,*

$$\|h_\gamma(x)\| \le C + 2\frac{\boldsymbol{d}(x)}{\gamma}\,,$$

*and moreover,*

$$\int (\|Y_\gamma(\xi, x)\|^2 + \|b(\xi, x)\|^2)^{\frac{1+\varepsilon}{2}} \mu(d\xi) \le C\left[1 + \left(\frac{\boldsymbol{d}(x)}{\gamma}\right)^{1+\varepsilon}\right].$$

*Proof.* Set $x \in K$, and introduce some $\tilde{x} \in \mathcal{D}$ such that $\|x - \tilde{x}\| \le 2\boldsymbol{d}(x)$. Relying on the fact that $A_\gamma(\xi, \,.\,)$ is $\frac{1}{\gamma}$-Lipschitz continuous,

$$\begin{aligned}
\|Y_\gamma(\xi, x)\| &\le \|A_\gamma(\xi, \tilde{x})\| + \frac{1}{\gamma}\|x - \gamma b(\xi, x) - \tilde{x}\| \\
&\le \|A_0(\xi, \tilde{x})\| + \|b(\xi, x)\| + 2\frac{\boldsymbol{d}(x)}{\gamma}\,.
\end{aligned}$$

Therefore,

$$\|h_\gamma(x)\| \le \int \|A_0(\xi,\tilde{x})\|\mu(d\xi) + 2\int \|b(\xi,x)\|\mu(d\xi) + 2\frac{\boldsymbol{d}(x)}{\gamma}\,.$$

The first two terms are independent of $\gamma$ and, by Assumptions 3 and 6, are bounded functions of $x$ on the compact $K$. This proves the first statement of the Lemma. Let $\varepsilon = \varepsilon(K)$ be the exponent defined in Assumption 3. There exists a constant $C$ such that

$$\begin{aligned}
(\|Y_\gamma(\xi,x)\|^2 + \|b(\xi,x)\|^2)^{\frac{1+\varepsilon}{2}} &\\
&\le C(\|Y_\gamma(\xi,x)\|^{1+\varepsilon} + \|b(\xi,x)\|^{1+\varepsilon})\\
&\le C\Big(\Big(\|A_0(\xi,\tilde{x})\| + \|b(\xi,x)\| + 2\frac{\boldsymbol{d}(x)}{\gamma}\Big)^{1+\varepsilon} + \|b(\xi,x)\|^{1+\varepsilon}\Big)\\
&\le C'\Big(2^\varepsilon\|A_0(\xi,\tilde{x})\|^{1+\varepsilon} + 2^{1+2\varepsilon}\|b(\xi,x)\|^{1+\varepsilon} + 2^{1+3\varepsilon}\Big(\frac{\boldsymbol{d}(x)}{\gamma}\Big)^{1+\varepsilon}\Big).
\end{aligned}$$

By Assumption 6 and since $\int\|b(\xi,x)\|^{1+\varepsilon}\mu(d\xi) \le 1+\int\|b(\xi,x)\|^2\mu(d\xi)$, there exists some (other) constant $C$ such that

$$\begin{aligned}
\int (\|Y_\gamma(\xi,x)\|^2 + \|b(\xi,x)\|^2)^{\frac{1+\varepsilon}{2}}\mu(d\xi) &\\
&\le C\Big(\int \|A_0(\xi,\tilde{x})\|^{1+\varepsilon}\mu(d\xi) + 1 + \|x\|^2 + \Big(\frac{\boldsymbol{d}(x)}{\gamma}\Big)^{1+\varepsilon}\Big).
\end{aligned}$$

The proof is concluded using Assumption 3. □

### End of the Proof of Theorem 3.1

Recall (10). Given an arbitrary real number $T > 0$, we shall study the asymptotic behavior of the family of functions $\{x(\tau_n+\cdot)\}_{n\in\mathbb{N}}$ on the compact interval $[0,T]$.

Given $\delta > 0$, we have $\|H(t+\delta) - H(t)\| \le \int_t^{t+\delta}\|h_{\gamma_{r(s)+1}}(x_{r(s)})\|ds$. By Proposition 6.1-1, the sequence $(x_n)$ is bounded a.s. Thus, by Lemma 6.3, there exists a constant $c_1 = c_1(\omega)$ such that for almost every $\omega$,

$$\begin{aligned}
\|H(t+\delta) - H(t)\| &\le c_1\delta + 2\int_t^{t+\delta}\frac{\boldsymbol{d}(x_{r(s)})}{\gamma_{r(s)+1}}ds\\
&\le c_1\delta + \int_t^{t+\delta}\left(1 + \frac{\boldsymbol{d}(x_{r(s)})^2}{\gamma_{r(s)+1}^2}\right)ds\\
&= (c_1+1)\delta + \int_t^{t+\delta}\frac{\boldsymbol{d}(x_{r(s)})^2}{\gamma_{r(s)+1}^2}ds\\
&\le (c_1+c_2+1)\delta + e(t)
\end{aligned}$$

for some $e(t) \to_{t\to\infty} 0$, where the last inequality is due to Proposition 6.2. We also observe from Proposition 6.1 and Assumption 6 that $M_n$ is a martingale in $\mathcal{L}^2(\Omega, \mathscr{F}, \mathbb{P}; \mathbb{R}^N)$, that

$$\mathbb{E}\|M_n\|^2 \le \mathbb{E}\Big[2\sum_{k=1}^{\infty} \gamma_k^2 \int \|Y_{\gamma_k}(\xi, x_k)\|^2 \mu(d\xi) + 2\sum_{k=1}^{\infty} \gamma_k^2 \int \|b(\xi, x_k)\|^2 \mu(d\xi)\Big] ,$$

and that the right-hand side is finite. Hence, $M_n$ converges almost surely. Therefore, on a probability one set, the family of continuous time processes $(M(\tau_n+\cdot)-M(\tau_n))_{n\in\mathbb{N}}$ converges to zero uniformly on $\mathbb{R}_+$. The consequence of these observations is that on a probability one set, the family of processes $\{z_n(\,.\,)\}_{n\in\mathbb{N}}$, where $z_n(t) = x(\tau_n+t)$, is equicontinuous. Specifically, for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\limsup_n \sup_{0\le t,s\le T, |t-s|\le\delta} \|z_n(t) - z_n(s)\| \le \varepsilon.$$

This family is moreover bounded by Proposition 6.1-1. By the Arzelà-Ascoli theorem, it has an accumulation point for the uniform convergence on $[0,T]$, for an arbitrary $T > 0$. From any sequence of integers, we can extract a subsequence (which we still denote as $(z_n)$ with slight abuse), and a continuous function $z(\cdot)$ on $[0,T]$, such that $(z_n)$ converges to $z$ uniformly on $[0,T]$. Hence, for $t \in [0,T]$,

$$\begin{aligned} z(t) - z(0) &= -\lim_{n\to\infty} \int_0^t h_{\gamma_{r(\tau_n+s)+1}}(x_{r(\tau_n+s)})\, ds \\ &= -\lim_{n\to\infty} \int_0^t ds \int_\Xi \mu(d\xi)\, (g_n^{(a)}(\xi,s) + g_n^{(b)}(\xi,s))\,, \end{aligned}$$

where we set $g_n^{(a)}(\xi,t) := Y_{\gamma_{r(\tau_n+s)+1}}(\xi, x_{r(\tau_n+s)})$ and $g_n^{(b)}(\xi,t) := b(\xi, x_{r(\tau_n+s)})$. Define the mapping $g_n := (g_n^{(a)}, g_n^{(b)})$ on $\Xi\times[0,T] \to \mathbb{R}^{2N}$. Recalling that the sequence $(\tilde{x}_n)$ belongs to a compact set, say $K$, let $\varepsilon \in ]0,1]$ be the exponent defined in Lemma 6.3. By the same Lemma,

$$\begin{aligned} \int_0^T ds \int_\Xi \mu(d\xi)\, \|g_n(\xi,s)\|^{1+\varepsilon} &\le c\Big[T + \int_0^T \Big(\frac{\boldsymbol{d}(x_{r(\tau_n+s)})}{\gamma_{r(\tau_n+s)+1}}\Big)^{1+\varepsilon} ds\Big] \\ &\le c\Big[T + T^{\frac{1-\varepsilon}{2}} \Big(\int_0^T \frac{\boldsymbol{d}(x_{r(\tau_n+s)})^2}{\gamma^2_{r(\tau_n+s)+1}} ds\Big)^{\frac{1+\varepsilon}{2}}\Big] \\ &\le c_1 \end{aligned}$$

for some constants $c$ and $c_1$. Therefore, the sequence of functions $(g_n)$ is bounded in $\mathcal{L}^{1+\varepsilon}(\Xi\times[0,T], \mathscr{T}\otimes\mathscr{B}([0,T]), \mu\otimes\lambda; \mathbb{R}^{2N})$, where $\lambda$ is the Lebesgue measure on $[0,T]$. The statement extends to the sequence of functions

$$\big(G_n(\xi,t) = (g_n(\xi,t), \|g_n^{(a)}(\xi,t)\|, \|g_n^{(b)}(\xi,t)\|)\big)_n,$$

which is uniformly bounded in $\mathcal{L}^{1+\varepsilon}(\Xi\times[0,T],\mathscr{T}\otimes\mathscr{B}([0,T]),\mu\otimes\lambda;\mathbb{R}^{2N+2})$. We can extract from this sequence a subsequence that converges weakly in this Banach space to a function $F:\Xi\times[0,T]\to\mathbb{R}^{2N+2}$. We decompose $F$ as $F(\xi,t)=(f(\xi,t),\kappa(\xi,t),\upsilon(\xi,t))$, where $\kappa,\upsilon$ are real-valued, and where $f(\xi,t)=(f^{(a)}(\xi,t),f^{(b)}(\xi,t))$ with $f^{(a)},f^{(b)}:\Xi\times[0,T]\to\mathbb{R}^N$. Using the weak convergence $(g_n^{(a)},g_n^{(b)})\rightharpoonup(f^{(a)},f^{(b)})$, we obtain

$$z(t)-z(0)=-\int_0^t ds\left(\int_\Xi f^{(a)}(\xi,s)\mu(d\xi)+\int_\Xi f^{(b)}(\xi,s)\mu(d\xi)\right).$$

It remains to prove that for almost every $t\in[0,T]$, $f^{(a)}(\,.\,,t)\in A(\,.\,,z(t))$ and $f^{(b)}(\,.\,,t)\in B(\,.\,,z(t))$ $\mu$-almost everywhere, along with $z(0)\in\mathrm{cl}(\mathcal{D})$. This shows that indeed $z(t)=\Phi(z(0),t)$ for every $t\in[0,T]$, and it follows that $x(t)$ is a.s. an APT of the differential inclusion (6).

By Mazur's theorem, there exists a function $J:\mathbb{N}\to\mathbb{N}$ and a sequence of sets of weights $(\{\alpha_{k,n},k=n\dots,J(n)\,:\,\alpha_{k,n}\geq 0,\sum_{k=n}^{J(n)}\alpha_{k,n}=1\})_n$ such that the sequence of functions defined by

$$\bar{G}_n(\xi,s)=\sum_{k=n}^{J(n)}\alpha_{k,n}\,G_k(\xi,s)$$

converges strongly to $F$. In the same way, we define $\bar{g}_n(\xi,s):=\sum_k\alpha_{k,n}\,g_k(\xi,s)$, and similarly for $\bar{g}_n^{(a)},\bar{g}_n^{(b)}$. Extracting a further subsequence, we obtain the $\mu\otimes\lambda$-almost everywhere convergence of $\bar{G}_n$ to $F$. By Fubini's theorem, for almost every $t\in[0,T]$, there exists a $\mu$-negligible set such that for every $\xi$ outside this set, $\bar{G}_n(\xi,t)\to F(\xi,t)$. From now on to the end of this proof, we fix such a $t\in[0,T]$.

As $\boldsymbol{d}(x_n)\to 0$, $z(t)\in\mathrm{cl}(\mathcal{D})$ (this holds in particular when $t=0$, hence $z(0)\in\mathrm{cl}(\mathcal{D})$). Following the same arguments as in the proof of Proposition 3.1, it holds that $z(t)\in\mathrm{cl}(D(\xi))$ for all $\xi$ outside a $\mu$-negligible set.

Define $\eta_n(\xi):=J_{\gamma_{m+1}}(\xi,x_m-\gamma_{m+1}b(\xi,x_m))-z(t)+\gamma_{m+1}b(\xi,x_m)$ with $m=r(\tau_n+t)$. Using the same approach as in the proof of Proposition 3.1, it can be shown that, as $n\to\infty$, $\eta_n(\,.\,)$ tends to zero almost surely along a subsequence. We now consider an arbitrary $\xi$ outside a $\mu$-negligible set, such that $\eta_n(\xi)\to 0$ and $z(t)\in\mathrm{cl}(D(\xi))$.

Let $(u,v)$ be an arbitrary element of $A(\xi,\cdot)$. By the monotonicity of $A(\xi,\cdot)$,

$$\langle v-Y_\gamma(\xi,x),u-J_\gamma(\xi,x-\gamma b(\xi,x))\rangle\geq 0\qquad(\forall x\in\mathbb{R}^N,\gamma>0),$$

and we obtain

$$\langle v - \bar{g}_n^{(a)}(\xi,t), u - z(t)\rangle = \sum_{k=n}^{J(n)} \alpha_{k,n} \langle v - g_k^{(a)}(\xi,t), u - z(t)\rangle$$

$$\geq \sum_{k=n}^{J(n)} \alpha_{k,n} \langle v - g_k^{(a)}(\xi,t), \eta_k(\xi,t) - \gamma_{r(\tau_k+t)+1} b(\xi, x_{r(\tau_k+t)})\rangle$$

$$\geq -\Big(\|v\| + \sum_{k=n}^{J(n)} \alpha_{k,n} \|g_k^{(a)}(\xi,t)\|\Big) \sup_{k\geq n} \big(\|\eta_k(\xi,t)\| + \gamma_{r(\tau_k+t)+1} \|b(\xi, x_{r(\tau_k+t)})\|\big)\,.$$

The term enclosed in the first parenthesis of the above right-hand side converges to $\|v\| + \kappa(\xi,t)$, while the supremum converges to zero using Assumption 6. As $\bar{g}_n^{(a)}(\xi,t) \to f^{(a)}(\xi,t)$, it follows that

$$\langle v - f^{(a)}(\xi,t), u - z(t)\rangle \geq 0\,,$$

and by the maximality of $A(\xi,\cdot)$, it holds that $f^{(a)}(\xi,t) \in A(\xi, z(t))$. The proof that $f^{(b)}(\xi,t) \in B(\xi, z(t))$ follows the same lines. □

### 6.3 Proof of Corollary 3.1

The proof is based on the study of the family of empirical measures of a process close to $x(t)$. Using [8], we show that any accumulation point of this family is an invariant measure for the flow $\Phi$. The corollary is then obtained by showing that the mean of such an invariant measure belongs to $\mathcal{Z}$.

Let $\boldsymbol{x}_n = \boldsymbol{\Pi}(x_n)$ be the projection of $x_n$ on $\mathrm{cl}(\mathcal{D})$, and write

$$\bar{\boldsymbol{x}}_n = \frac{\sum_{k=1}^n \gamma_k \boldsymbol{x}_k}{\sum_{k=1}^n \gamma_k}.$$

Let $\boldsymbol{x}(\omega,t)$ be the $\Omega \times \mathbb{R}_+ \to \mathbb{R}^N$ process obtained from the piecewise constant interpolation of the sequence $(\boldsymbol{x}_n)$, namely $\boldsymbol{x}(\omega,t) = \boldsymbol{x}_n$ for $t \in [\tau_n, \tau_{n+1}[$. On $(\Omega, \mathscr{F}, \mathbb{P})$, let $(\mathscr{F}_t)$ be the filtration generated by the process obtained from the similar piecewise constant interpolation of $(u_n)$. With regard to this filtration, $\boldsymbol{x}$ is progressively measurable. It is moreover obvious that $\boldsymbol{x}(\omega,\cdot)$ is an APT for (6) for almost all values of $\omega$. Let $\{\nu_t(\omega,\cdot)\}_{t\geq 0}$ be the family of empirical measures of $\boldsymbol{x}(\omega,\cdot)$. Observe from Theorem 3.1 that for almost all $\omega$, there is a compact set $K(\omega)$ such that the support $\mathrm{supp}(\nu_t(\omega,\cdot))$ is included in $K(\omega)$ for all $t \geq 0$, which shows that the family $\{\nu_t(\omega,\cdot)\}_{t\geq 0}$ is tight. Hence this family has accumulation points. Let $\nu$ be the weak limit of $(\nu_{t_n})$ along some sequence $(t_n)$ of times. By [8, Th. 1], $\nu$ is invariant for the flow $\Phi$. Clearly, $\mathrm{supp}(\nu)$ is a compact subset of $\mathrm{cl}(\mathcal{D})$. Moreover, for any $x \in \mathrm{supp}(\nu)$ and any $t \geq 0$, $\Phi(x,t) \in \mathrm{supp}(\nu)$. Indeed, suppose for the sake of contradiction that there exists $t_0 > 0$ such that

$\Phi(x, t_0) \notin \operatorname{supp}(\nu)$. Then, $\Phi(B(x,\varepsilon) \cap \operatorname{cl}(\mathcal{D}), t_0) \subset \operatorname{supp}(\nu)^{\mathrm{c}}$ for some $\varepsilon > 0$ by the continuity of $\Phi$ and the closedness of $\operatorname{supp}(\nu)$, where $B(x,\varepsilon)$ is the closed ball with centre $x$ and radius $\varepsilon$. Since $\nu(\Phi(B(x,\varepsilon)\cap\operatorname{cl}(\mathcal{D}), 0)) > 0$, we obtain a contradiction. We also know from [42] or [20, Th. 5.3] that there exists $\varphi : \operatorname{cl}(\mathcal{D}) \to \mathcal{Z}$ such that

$$\forall x \in \operatorname{cl}(\mathcal{D}), \quad \frac{1}{t}\int_0^t \Phi(x,s)\, ds \xrightarrow[t\to\infty]{} \varphi(x).$$

By the dominated convergence and Fubini's theorems, we now have

$$\begin{aligned}\int \varphi(x)\, \nu(dx) &= \int \nu(dx) \lim_{t\to\infty} \frac{1}{t}\int_0^t ds\, \Phi(x,s) = \lim_{t\to\infty} \frac{1}{t}\int_0^t ds \int \nu(dx)\, \Phi(x,s) \\ &= \int x\, \nu(dx)\,,\end{aligned}$$

which shows that $\int x\, \nu(dx) \in \mathcal{Z}$ by the convexity of this set. Since we have $\int x\, d\nu_{t_n} \to \int x\, d\nu$ as $n \to \infty$, we conclude that all the accumulation points of $(\bar{\boldsymbol{x}}_n)$ belong to $\mathcal{Z}$. On the other hand, since $\mathcal{R}_{2p}(x_\star) \neq \emptyset$ for each $x_\star \in \mathcal{Z}$, a straightforward inspection of the proof of Proposition 6.1-3. shows that $(\|x_n - x_\star\|)$ converges almost surely for each $x_\star \in \mathcal{Z}$. From these two facts, we obtain by [32] or [20, Lm 4.2] that $(\bar{\boldsymbol{x}}_n)$ converges a.s. to a point of $\mathcal{Z}$. Since $x_n - \boldsymbol{x}_n \to 0$ a.s., the convergence of $(\bar{x}_n)$ to the same point follows. $\square$

### 6.4 Proof of Corollary 3.2

Let us start with a preliminary lemma.

**Lemma 6.4.** *Let* $\mathsf{A} \in \mathcal{M}$ *be demipositive. Assume that the set* $\operatorname{zer}(\mathsf{A})$ *of zeros of* $\mathsf{A}$ *is not empty. Let* $\Psi : \operatorname{cl}(\operatorname{dom}(\mathsf{A})) \times \mathbb{R}_+ \to \operatorname{cl}(\operatorname{dom}(\mathsf{A}))$ *be the semiflow associated to the differential inclusion* $\dot{z}(t) \in -\mathsf{A}(z(t))$. *Then, any ICT set of* $\Psi$ *is included in* $\operatorname{zer}(\mathsf{A})$.

*Proof.* Let $K$ be an ICT set and let $U$ be an arbitrary, bounded and open set of $\mathbb{R}^N$ such that $K \cap U \neq \emptyset$. Define $G_t := \bigcup_{s\geq t} \Psi(U, s)$ for all $t \geq 0$. For any $x_* \in \operatorname{zer}(\mathsf{A})$ and any $x \in U$,

$$\|\Psi(x,t)\| \leq \|\Psi(x,t) - \Psi(x_*,t)\| + \|x_*\| \leq \|x - x_*\| + \|x_*\|\,.$$

Therefore, $G_0$ is a bounded set. By [4, Prop. 3.10], the set $G = \bigcap_{t\geq 0} \operatorname{cl}(G_t)$ is an attractor for $\Psi$ with a fundamental neighbourhood $U$. As $K \cap U \neq \emptyset$, it follows that $K \subset G$ by [17, Corollary 5.4]. We finally check that $G \subset \operatorname{zer}(\mathsf{A})$. Let $y \in G$, that is, $y = \lim_{k\to\infty} \Psi(x_k, t_k)$ for some sequence $(x_k, t_k)$ such that $x_k \in U$ and $t_k \to \infty$. By compactness of $\operatorname{cl}(U)$, the sequence $x_k$ can be chosen such that $x_k \to \bar{x}$ for some $\bar{x} \in \operatorname{cl}(U)$. Therefore, $y = \lim_{k\to\infty} \Psi(\bar{x}, t_k)$, which by demipositivity of $\mathsf{A}$, implies $y \in \operatorname{zer}(\mathsf{A})$ [9, 20]. $\square$

By Theorem 3.1 and the discussion of Section 2.4, $L(x)$ is an ICT set. Using Lemma 6.4 and the standing hypotheses, $L(x) \subset \mathcal{Z}$. On the other hand, since $\mathcal{R}_2(x_\star) \neq \emptyset$ for all $x_* \in \mathcal{Z}$, a straightforward inspection of the proof of Proposition 6.1-3. shows that $\|x_n - x_*\|$ converges almost surely for any of those $x_*$. By Opial's lemma [20, Lm 4.1], we obtain the almost sure convergence of $(x_n)$ to a point of $\mathcal{Z}$. □

### 6.5 Proof of Corollary 4.1

Define the probability distribution $\zeta := \sum_{i=0}^m \alpha_i \delta_i$ on $\{0, 1, \dots, m\}$.

On the space $\mathsf{X} \times \{0, \dots, m\}$ equipped with the probability $\mu = \nu \otimes \zeta$, let $\xi = (\eta, i)$, and define the random operators $A$ and $B$ by

$$A(\xi, \cdot) := \begin{cases} \alpha_0^{-1} \partial_x g(\eta, \cdot), & \text{if } i = 0, \\ N_{\mathcal{C}_i}, & \text{otherwise}, \end{cases} \quad \text{and} \quad B(\xi, \cdot) := \partial_x f(\eta, \cdot).$$

The Aumann integral $\mathcal{B}(x) = \int \partial f(\eta, x) d\pi(\eta)$ coincides with $\partial F(x)$ by [43] (see also the discussion in Section 4.1). Similarly, $\mathcal{A}(x) = \partial(G(x) + \iota_{\mathcal{C}})(x)$. The operator $\mathcal{A}$ is thus maximal. It holds that $\mathcal{A} + \mathcal{B} = \partial(F + G + \iota_{\mathcal{C}})$, which is maximal, demipositive, and whose zeros coincide with the minimizers of $F + G$ over $\mathcal{C}$. The end of the proof consists in checking the assumptions of Corollary 3.2. It follows the same line as [7] and is left to the reader. □

## 7 Perspectives

Beyond the forward-backward algorithm, the concept of random maximal monotone operators can be used to study stochastic versions of other popular optimization algorithms that rely on the monotone operator theory. Our next research direction is therefore to extend our approach to other kinds of algorithms, such as the Douglas-Rachford algorithm, as a way to construct new families of stochastic approximation algorithms. In this perspective, the present paper may contain useful ingredients.

It would also be interesting to weaken the assumption that the "innovation" $(u_n)$ is an iid sequence. More involved random models are often useful. Among those are the ones where the innovation is a Markov chain controlled by the iterates. Such models are popular in the classical stochastic approximation literature.

Another research direction includes the case where the step size of the algorithm is constant. In this context, the APT property does not hold and the iterates are no longer expected to converge a.s., due to the persistence of the random effects. Tools from the weak convergence theory of stochastic processes can be useful to address this setting.

Finally, we believe that our algorithm can be shown to be useful to address several specific applications in the field of convex optimization and

variational inequalities. An important aspect is to instanciate the algorithm in practical scenarios related to machine learning, signal processing, or game theory.

# 8 Conclusions

The question of providing stochastic versions of well-known deterministic algorithms relying on maximal monotone operators has become increasingly popular. In particular, several authors have studied the effects of additive random errors on the behavior of the iterates, showing that the errors have no effect on the limiting points, provided some adequate vanishing condition of the former. The approach taken by this paper is conceptually different in the sense that the operators themselves are assumed to be random. This situation involves two key-ingredients. The first one is the Aumann expectation of the random operators. The second one is the notion of asymptotic pseudotrajectory, borrowed from Benaïm and Hirsch, which is used to relate the iterates to a continuous-time dynamical system.

# Acknowledgements

This work was partially funded by phi-TAB, the Orange - Telecom ParisTech think tank, and by the ASTRID program of the French Agence Nationale de la Recherche (ODISSEE project ANR-13-ASTR-0030).

# References

[1] Mercier, B.: Lectures on topics in finite element solution of elliptic problems, Tata Institute of Fundamental Research Lectures on Mathematics and Physics, vol. 63. Tata Institute of Fundamental Research, Bombay (1979). With notes by G. Vijayasundaram

[2] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)

[3] Benaïm, M., Hirsch, M.W.: Asymptotic pseudotrajectories and chain recurrent flows, with applications. J. Dynam. Differential Equations **8**(1), 141–176 (1996)

[4] Benaïm, M., Hofbauer, J., Sorin, S.: Stochastic approximations and differential inclusions. SIAM J. Control Optim. **44**(1), 328–348 (electronic) (2005)

[5] Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. **12**, 1–12 (1965)

[6] Aubin, J.P., Frankowska, H.: Set-valued analysis. Modern Birkhäuser Classics. Birkhäuser Boston, Inc., Boston, MA (2009). Reprint of the 1990 edition

[7] Bianchi, P.: Ergodic convergence of a stochastic proximal point algorithm. ArXiv e-prints, 1504.05400 (2015)

[8] Benaïm, M., Schreiber, S.J.: Ergodic properties of weak asymptotic pseudotrajectories for semiflows. J. Dynam. Differential Equations **12**(3), 579–598 (2000)

[9] Bruck Jr., R.E.: Asymptotic convergence of nonlinear contraction semigroups in Hilbert space. J. Funct. Anal. **18**, 15–26 (1975)

[10] Brézis, H.: Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland mathematics studies. Elsevier Science, Burlington, MA (1973)

[11] Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York (2011)

[12] Castaing, C., Valadier, M.: Convex analysis and measurable multifunctions, vol. 580. Springer (1977)

[13] Hiai, F., Umegaki, H.: Integrals, conditional expectations, and martingales of multivalued functions. Journal of Multivariate Analysis **7**(1), 149 – 182 (1977)

[14] Attouch, H.: Familles d'opérateurs maximaux monotones et mesurabilité. Annali di Matematica Pura ed Applicata **120**(1), 35–111 (1979). DOI 10.1007/BF02411939

[15] Hiriart-Urruty, J.B.: Contributions à la programmation mathématique: cas déterministe et stochastique. Université de Clermont-Ferrand II, Clermont-Ferrand (1977). Thèse présentée à l'Université de Clermont-Ferrand II pour obtenir le grade de Docteur ès Sciences Mathématiques, Série E, No. 247

[16] Aubin, J.P., Cellina, A.: Differential inclusions, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 264. Springer-Verlag, Berlin (1984). Set-valued maps and viability theory

[17] Benaïm, M.: Dynamics of stochastic approximation algorithms. In: Séminaire de Probabilités, XXXIII, Lecture Notes in Math., vol. 1709, pp. 1–68. Springer, Berlin (1999)

[18] Delyon, B.: Stochastic Approximation with Decreasing Gain: Convergence and Asymptotic Theory. Unpublished Lecture Notes, http://perso.univ-rennes1.fr/bernard.delyon/as_cours.ps (2000)

[19] Bauschke, H.H., Borwein, J.M., Li, W.: Strong conical hull intersection property, bounded linear regularity, jamesons property (g), and error bounds in convex optimization. Mathematical Programming **86**(1), 135–160 (1999)

[20] Peypouquet, J., Sorin, S.: Evolution equations for maximal monotone operators: asymptotic analysis in continuous and discrete time. J. Convex Anal. **17**(3-4), 1113–1163 (2010)

[21] Pazy, A.: On the asymptotic behavior of semigroups of nonlinear contractions in Hilbert space. J. Functional Analysis **27**(3), 292–307 (1978)

[22] Rockafellar, R.T.: Measurable dependence of convex sets and functions on parameters. J. Math. Anal. Appl. **28**, 4–25 (1969)

[23] Rockafellar, R.T., Wets, R.J.B.: Variational analysis, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317. Springer-Verlag, Berlin (1998)

[24] Walkup, D.W., Wets, R.J.B.: Stochastic programs with recourse. II: On the continuity of the objective. SIAM J. Appl. Math. **17**, 98–103 (1969)

[25] Kushner, H.J., Yin, G.G.: Stochastic approximation and recursive algorithms and applications, Applications of Mathematics (New York), vol. 35, second edn. Springer-Verlag, New York (2003). Stochastic Modelling and Applied Probability

[26] Nedić, A.: Random algorithms for convex minimization problems. Math. Program. **129**(2, Ser. B), 225–253 (2011)

[27] Lee, S., Nedic, A.: Distributed random projection algorithm for convex optimization. Selected Topics in Signal Processing, IEEE Journal of **7**(2), 221–229 (2013). DOI 10.1109/JSTSP.2013.2247023

[28] Bertsekas, D.P.: Incremental proximal methods for large scale convex optimization. Math. Program. **129**(2, Ser. B), 163–195 (2011)

[29] Wang, M., Bertsekas, D.P.: Incremental constraint projection-proximal methods for nonsmooth convex optimization. Tech. rep., Massachusetts Institute of Technology (2013)

[30] Wang, M., Bertsekas, D.P.: Incremental constraint projection methods for variational inequalities. Math. Program. **150**(2, Ser. A), 321–363 (2015)

[31] Kinderlehrer, D., Stampacchia, G.: An introduction to variational inequalities and their applications, Classics in Applied Mathematics, vol. 31. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2000). Reprint of the 1980 original

[32] Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**(2), 383–390 (1979)

[33] Combettes, P., Pesquet, J.C.: Stochastic approximations and perturbations in forward-backward splitting for monotone operators. Pure and Applied Functional Analysis (2016). To appear

[34] Atchade, Y.F., Fort, G., Moulines, E.: On stochastic proximal gradient algorithms. ArXiv e-prints, 1402.2365 (2014)

[35] Rosasco, L., Villa, S., Vũ, B.C.: Convergence of stochastic proximal gradient algorithm. arXiv preprint arXiv:1403.5074 (2014)

[36] Rosasco, L., Villa, S., Vũ, B.C.: A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions. arXiv preprint arXiv:1507.00848 (2015)

[37] Toulis, P., Tran, D., Airoldi, E.M.: Stability and optimality in stochastic gradient descent. ArXiv e-prints, 1505.02417 (2015)

[38] Alvarez, F., Peypouquet, J.: Asymptotic equivalence and Kobayashi-type estimates for nonautonomous monotone operators in Banach spaces. Discrete Contin. Dyn. Syst. **25**(4), 1109–1128 (2009)

[39] Alvarez, F., Peypouquet, J.: Asymptotic almost-equivalence of Lipschitz evolution systems in Banach spaces. Nonlinear Analysis: Theory, Methods & Applications **73**(9), 3018–3033 (2010)

[40] Álvarez, F., Peypouquet, J.: A unified approach to the asymptotic almost-equivalence of evolution systems without Lipschitz conditions. Nonlinear Anal. **74**(11), 3440–3444 (2011)

[41] Robbins, H., Siegmund, D.: A convergence theorem for non negative almost supermartingales and some applications. In: Optimizing Methods in Statistics, pp. 233–257. Academic Press, New York (1971)

[42] Baillon, J.B., Brézis, H.: Une remarque sur le comportement asymptotique des semigroupes non linéaires. Houston J. Math. **2**(1), 5–7 (1976)

[43] Rockafellar, R.T., Wets, R.J.B.: On the interchange of subdifferentiation and conditional expectations for convex functionals. Stochastics **7**(3), 173–182 (1982)